

From: Chan, Christina
Sent: Monday, February 11, 2002 4:06 PM
To: Davis, Minh-Tam; STIC-Biotech/ChemLib
Subject: RE: Rush search request for 09/775693

Please rush. Thanks Chris

-----Original Message-----

From: Davis, Minh-Tam
Sent: Monday, February 11, 2002 3:34 PM
To: Chan, Christina
Subject: Rush search request for 09/775693

Please search in commercial database and in issued patent files:

1) SEQ ID NO:1 and a polypeptide encoded by SEQ ID NO:1. Please check to see any of said polypeptide is the same as SEQ ID NO:7 or 8.

Thank you.

MINH TAM DAVIS
ART UNIT 1642, ROOM 8A01, MB 8E12
305-2008

RECEIVED
FEB 11 2002
STIC

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 Tel: 305-4994
1E05
508-4994

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

SEARCH REQUEST FORM

Requestor's
Name: _____Serial
Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 02-13-02Searcher: Beverly 4994Terminal time: 20

Elapsed time: _____

CPU time: _____

Total time: 25

Number of Searches: _____

Number of Databases: 1

Search Site

STIC

CM-1

Pre-S

Type of Search

N.A. Sequence

A.A. Sequence

Structure

Bibliographic

Vendors

IG

STN

Dialog

APS

Geninfo

SDC

DARC/Questel

☒ Other

CGN


```
seq_name: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT: AAB57016
seq_documentation_block: Protein: 442 AA.
ID AAB57016 standard;
XX
XX AAB57016;
```

[illegible][illegible]

1 ATGTCGACGCAAAGGCTGGG

314 hrrProAlaGlyThrIleLeuValrHISAlaHisLeuAspIleGluAlaPhe 330
901 ACCATGACCGGAGAGTGCGCAAAATCAAAACAAGCGCGGCTTGAATT 950
331 ThirHeAspArgGluValaLgLySileTySIngInIleuLeuTySph 950
951 TGCTGACGTGGTATACCGGCTTACGCGCTAGCGCTGAGCTGAATTG 347
347 eAlaGluLeuValrTyThrGlyPheThrPHisSerProGluCysGluPheV 1000
1001 TCCGCACTGCATTCGCGCAAGTCCGAGGAGGAGTGGAAAGGAAAGTCAG 1050
1051 GTGCGGTCTTCAGAGGCGCAGGTGTACATCCGCGCGGAGTCCCACT 1100
381 ValSerValIleuTySIngInValrTyThrIleuLeuValrGluTySph 380
1101 GTCTCTACATATGAGACCTGTGACATGACGAGCAAGCGAGGTATATG 397
397 uSerLeuTyHAsnGluGluLeuValSerMetAsnValGInIAspTyG 414
1151 AGCCAACTGATCCACCGGGTTCATCATCAACATCAATTCCTCAGCTGAAG 1200
414 IupProHisPheIaThrGlyPheIleAsnIleAsnSerLeuArgIleuTyS 430
1201 GAATTCATCGTCTCCAGCAAGGCTACTGCCAA 1236
431 GluTyHISArgIleuInserrTySValrHrAlaTyS 442
seq.name: /STDS2/gcdata/geneseq/geneseqp/AA2001.DAT.AAG75416
seq_documentation_block:
ID AAG75416 standard; Protein: 442 AA.
AC AAG75416;
XX
DT 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SpQ ID NO:6180.
KW Human: colon cancer; colon cancer antigen; diagnosis; detection.
XX colorectal carcinoma.
XX Homo sapiens.
OS
XX
XX Homo sapiens.
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
XX
PR 03-NOV-1999; 99US-0163280.
(HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Blrse CE, Rosen CA;
XX WPT: 2001-235357/24.
XX N-PSDB: AAH34821.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11: Page 7623-7624; 9803BP; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as nucleic acid molecules (N) and
XX therapy antigens have cytostatic activity; colon cancer anti-
XX diant and vaccine nucleic acid molecules (N) and proteins (P), where

[illegible]

214 snleumethistilesertylgluablagllyleuengluasnprolylsasn 230
601 CAAGCGCTCTCAGGTCCTTACAGCAGAACCCAGACCCAGCCAAAGCCCC 650
|||||
231 glialleproproglyleutythrlysthrghlasprralibylalapr 247
651 CACACCCCTGCACATTTCTCGAGATCGAGTTTCAAAAAGGGGTCCCTGGA 700
oasrthrpriasprialeuenglulleghlyubhelyslsglyalvalprovalr 264
247 oasrthrpriasprialeuenglulleghlyubhelyslsglyalvalprovalr 264
701 AGGTGACCAACGTCAGAGATGCGCACACACCCACGAGCTCTTGGAGCTC 750
264 yslvalthrlnasvalylsasprglythrthrhisglnthrserleuengluleu 280
751 TTTCATGATACCTGAACGAGTCGGCGGCAAGCATGGCGGTGGCGGTATTGA 800
|||||
281 pheketylrlleuaengluvalalaglylsthlsghlyalaglyalglleas 297
801 CATGTGGAGAACCCCTTCATTTGGATGAAGTCCCGAGATGCTATCAGAGA 850
297 rllvalagluasnhrghleuethlysmetylserhrnglylletyrglutr 314
851 CCCGACGACGACACATCCCTTACCATGCTCATTTAGACATCGAGGCTTC 900
314 hrprohlaelythrleuetyrhlsalhisleuasprialeuabhe 330
901 ACCATGAGCCGGGAGTCCGGAATGAATGAACAGGCTGGGCTTGAATT 950
331 thmetaspargelubalargylsllelysglnlyleuenglyleuylsph 347
951 TGTGAGCTGGTGTATACCGGTTTACGGCTTACGGCTCGAGTGTGATTGTG 1000
347 ealaglualeuvaltyrthglythrphlserproglucylglubhe 364
1001 TCGGCACTGCATCGCCAGATCCGAGAGCAGATGGAAGGAAGAGCAG 1050
364 alarghiscysllealalyserglnghuagvalghuglylvalgln 380
1051 GTGTCCGTCTCAAGGGCCAGGTGATCATCTCGGCCGGAGTCCCACT 1100
381 valservalleuylslgylgluvaltyrilleuenglylartgylserprole 397
1101 GTCCTCTACATGAGAGAGCTGTGAGCATGGAACGTCGAGGTGATTATG 1150
397 userleutyraengluuglualeuvalsermekasnvalglnghlyasprtyr 414
1151 AGCCACATGATGCCACCGGGTTTCATCAACATCAATTCCCTCAGCGTGAG 1200
414 luprothrasplalathrnglypheellesnllleasnserleuargyleuyls 430
1201 GAATATCATGCTCTCCAGCAGCAGAGGTCTACTGCCAAA 1236
|||||
431 glutyrlhisargleuulserlyslvalthrhalays 442
seq_name: /STDs2/gcdata/geneseq/geneseq/AA2001.DAT:AA682054
seq_documentation_block:
ID AA682054 standard: Protein: 367 AA.
AC AA682054:
XX
AC
XX 03-SEP-2001 (first entry)
DT
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:1202
XX
XX Staphylococcus epidermidis SRI strain: infection: diagnosis;
KW vaccination; endocarditis.
XX
XX Staphylococcus epidermidis.
OS
XX
XX WO200134809-A2.
PN
XX
17-MAY-2001.
PD

CC (II), given in AAG81453, represent nucleic acids (I) encoding polypeptides
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria
CC activity and therefore may also be used to assay for other inhibitors of their
CC treatment of S. epidermidis infections, e.g., endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO.:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO.:4455 to 4472,
XX no sequences are present for SEQ ID NO.:4455 to 4464.
XZ Sequence 367 AA:

Align seg 1/1 to: AAG82054 from: 1 to: 367

[illegible]

```

85 yrProLeuValSerAlaLeuSerArpProLeuIleAlaLysLysLeuVal 101
310 GAAATCCGCCAGCGGGAGGGGCCAAATATGTGTCCACGCGCCACAGG 359
102 GluIleAlaGluLysrThrAnserIleGlyIleAlaHisGlyLysrThr 118
360 AAAGGGGAAGCATCAGAGTCCGGTTTGAGCTCAGACTCTACTCTACGGCC 409
118 yurSgLYsAnsrGlnValArGpHeLyuValAlaIleLysAlaLeuAnsr 135
410 CCAGATAAAGGTCAATTGCTCCCTGGAGGATCGCTCAATTCTACAACCGG 459
135 rOlYsLeuLysAlaPheAlaProValArGluLTrPalaTrP..... 148
460 TTCAGGGCCGCAATGACCTGTAGATACCGAAGCAACACGGGATTCC 509
149 .....SerArGgLuGluGluIleAspYrAlaIleLysHisAsnIlePr 163
510 CATCCCGGTCACTCCCAAGAACCCGTGGAGCATGAGTGAAGAACTCATGC 559
163 oValSerIleAsnYrAsrProYrSerIleAspHisAsnLeuTrpG 180
560 ACATCAGCTTCAGAGCTGTGAATCTGTGAGAACCCCAAGAACACGGCTT 609
180 LyArAlaAsnGluCysGlyIleLeuLysProYrAlaIleArPro 196
610 CCAAGCTCTACACGAAGACCCAGGACCCGCAACCCCAACACCC 659
197 GluAspAlaPheAsrLeuThrTrProLeuGluGluTrProAspSnaI 213
660 TGCATTTCTCGAGTGCAGTCAAAAAGGGCTCCGTGAAGGTGACCA 709
213 aAspGluIleIleLeuThrPheLysSerArGluIleYrGluTrProGlyA 228
710 AGCTAAGAGTGGACACCCACACACCTCTGGAGCTCTTCAATGTAC 759
229 .....AspLYsAsrYrGluGlnLeuAsp...AspLeuIleLeuTr 241
760 CTGAACGAGTTCGGCGGCAACGATGCGTGGCGGTATTCACATGCTGA 809
242 LeuAsnGlnIleuAlaGlyLysGlyIleuYrGlyLeAspHisValGI 258
810 GAACGCGTTCATGGAATGAAGTCCGAGATATCTAGACACCCACAGC 859
258 uAsnArGmetValoLYrIleLysSerArGluIleYrGluTrProGlyA 275
860 GCACACATCTTACATGCTCATTTAGACATCTAGAGGCGTCCACATGAC 909
275 IAcLYuAlIleLysLysAlaHisLYuAlaLeuGluTrIleTrLeuTr 921
910 CGGAGTGCAGCAAAATCAACACAGGCTCGGCTTGAATTTGCTAGCT 959
292 LysAspValAlaHisPheLysProValIleGluLYsGlnPheSerGluI 308
960 GGTGTATACCGGTTTACCGCTAGCGCTGAGCTGATAATTGCCCCACT 1009
308 nIleLYrAsnGlyLeuTrPheSerProLeuThrAspSerLeuLysLeuP 325
1010 GCATGCCCAAGTCCACGAGCGAGTGAAGGAAGAAATGTCACAGTGTCTC 1059
325 heIleAspSerThrGlnoLYrValIGluLYsAspValArGIleLysLeu 341
1060 CTCAGAGGCCAGGTGTACATCTTCGGCGGAGGAGCCACAGCTGTCTTA 1109
342 PheLYsLYsAlaAlaIleValaLAsnGlyArGlnSerProLYrThrLeuTr 1150
1110 CAATGAGAGACTGTGAGC 1128
1110 rAspGluLYsLeuAlaThr 364

seq_name: /SID2/gcgdata/geneseq/geneseqp/AA2001.DAT:AA691292
seq_documentation_block:

```



```

322 erLeuAspAlaPheIleAspSerThrGlnGluHisValThrGlyAspIle 338
1048 CAGGTGTCCTGCTCAAGGCCGAGGTGTCATCTCCGCCGAGATCCCC 1097
339 ArgMetValLeuHisIleGlySerIleThrIleAsnGlyArgArgSerSe 355
1098 ACCTGCTCTTCACATGAGAGCTGTCAGTCATGACCTGAGGTCATT 1147
355 rHisSerLeuTyrAspPheAsnLeuAlaThrTyrAspThrGlyAspThr 372
1148 ATGAGCCACTGATGACCGGTTTCATCAACATTCCTCAGGCTG 1197
372 heAspIleThrLeuAlaIleGlyPheValGlnLeu..... 383
1198 AAGGATATCATCTCCAGACAGCTCACATCCCAAA 1236
384 .....HisGlyLeuSerSerLysIleAlaAsnLys 393

seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT: AAB79801
seq_documentation_block:
ID AAB79801 standard; Protein: 401 AA.
AC AAB79801;
XX
XX 30-APR-2001 (first entry)
XX
XX Corynebacterium glutamicum MP protein sequence SPQ ID NO:336.
XX
XX fine chemical production; metabolic pathway protein; MP protein;
XX nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
XX lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
XX carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX
XX Corynebacterium glutamicum.
XX
XX MO200100843-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-IB00923.
XX
XX 25-JUN-1999; 99US-0141031.
XX 01-JUL-1999; 99DE-1030476.
XX 02-JUL-1999; 99US-0142101.
XX 08-JUL-1999; 99DE-1031415.
XX 08-JUL-1999; 99DE-1031418.
XX 08-JUL-1999; 99DE-1031419.
XX 08-JUL-1999; 99DE-1031420.
XX 08-JUL-1999; 99DE-1031424.
XX 08-JUL-1999; 99DE-1031428.
XX 08-JUL-1999; 99DE-1031434.
XX 08-JUL-1999; 99DE-1031435.
XX 08-JUL-1999; 99DE-1031443.
XX 08-JUL-1999; 99DE-1031453.
XX 08-JUL-1999; 99DE-1031457.
XX 08-JUL-1999; 99DE-1031478.
XX 08-JUL-1999; 99DE-1031510.
XX 08-JUL-1999; 99DE-1031541.
XX 08-JUL-1999; 99DE-1031592.
XX 08-JUL-1999; 99DE-1031632.
XX 08-JUL-1999; 99DE-1031634.
XX 08-JUL-1999; 99DE-1031636.
XX 09-JUL-1999; 99DE-1032125.
XX 09-JUL-1999; 99DE-1032126.
XX 09-JUL-1999; 99DE-1032130.
XX 09-JUL-1999; 99DE-1032186.
XX 09-JUL-1999; 99DE-1032206.
XX 09-JUL-1999; 99DE-1032227.

```

```

PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 31-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041386.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.

(PADI ) BASF AG.
XX
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX N-PSDB; AAF71920.
XX
XX WPI: 2001-137957/14.
XX
XX DR N-PSDB; AAF71920.
XX
XX PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
XX PT pathway proteins, useful for producing fine chemicals in
XX PT microorganisms, including organic acids, nonproteinogenic amino
XX PT and purine and pyrimidine bases -
XX
XX Claim 20; Page 656-657; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
XX CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
XX CC MP nucleic acids are useful for the production of fine chemicals
XX CC in microorganisms, including organic acids, nonproteinogenic amino
XX CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
XX CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
XX CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
XX Sequence 401 AA;

```

alignment_scores:

Quality:	619.50	Length:	413
Ratio:	2.261	Gaps:	8
Percent Similarity:	66.344	Percent Identity:	35.351

alignment_block:

US-09-775-693-1 x AAB79801 ..

Align seg 1/1. to: AAB79801 from: 1 to: 401

```

19 GTGGTTCGCTACAGATGAGCGGCTGACACCTGCTCATCTCGCTG 68
   : : : : : : : : : : : : : : : : : : : : : : : : : :
5  IleValLeuAlaTyrSerGlyLeuAspThrValAlaIleProTy 21
   : : : : : : : : : : : : : : : : : : : : : : : : : :
69 GCCTATGACGTCATTGCCATATCTGCGCAACATG 115
   : : : : : : : : : : : : : : : : : : : : : : : : : :
21 rLeuLysTyrMetIleAspGlyIleValIleAlaValSerLeuAspLeu 38

```


PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 24-JUN-1999; 99US-0140351.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142380.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 11-AUG-1999; 99US-0147935.
PR 12-AUG-1999; 99US-0148171.
PR 13-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151348.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154038.
PR 20-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156559.
PR 29-SEP-1999; 99US-0156458.
PR 05-OCT-1999; 99US-0156596.
PR 06-OCT-1999; 99US-0157717.
PR 07-OCT-1999; 99US-0157855.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

28-0CT-1999; 99US-0161992.
 PR 28-0CT-1999; 99US-0161993.
 PR 29-0CT-1999; 99US-0162142.

alignment_scores:

Quality: 533.00 Length: 203
 Ratio: 3.270 Gaps: 3
 Percent Similarity: 80.296 Percent Identity: 50.246

alignment_block:

US-09-775-693-1 x AAC08821 ..

Align seg 1/1 to: AAC08821 from: 1 to: 302

```

19 GTGCTTGGCCAGTGGAGGCGCTGAGACCTGCTGCTGCTGCTG 68
   |||||
99 ValValLeuAlaTyrSerGlyLeuAspThrSerValIleValProtr 115
69 GCTGAAGAACAA...GGCTATGACGTCTATGCTGCTGCTGCTG 115
   |||||
115 PleuLeuGluAsnTyrGlyCysGluValValCysPheThrAlaAspValG 132
   |||||
116 GCCGGAAG...GAAGACTTGAGAGAACCCAGAAAGGACGACTGA 162
   |||||
132 LysGlyIleLeuGluLeuGluLeuGluLeuGluLeuGluLeuGlu 148
   |||||
163 GGGGCAAAAGGTTCATTGAGATGTCAGCAGGAGGTTTGTGAGCA 212
   |||||
149 GlyAlaSerGluLeuValAlaAsnAspLeuThrGluLeuValLysAs 165
   |||||
213 GTTCATCTGGCCGGCCATCCAGTCCAGCGACGTGATGAGCGCTAC 262
   |||||
165 PheIlePheProCysLeuArgAlaGlyAlaIleTyrGluArgLysTyr 182
   |||||
263 TCCCTGGGACCTCTCTGTCAGAGCCCTGTCATGCCCGCAACAA 312
   |||||
182 euLeuGlyThrSerMetAlaArgProValIleAlaLysAlaMetValAs 198
   |||||
313 ATCCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 362
   |||||
199 ValAlaIleGluValGlyAlaAspAlaValAlaIleGlyCysThrGly 215
   |||||
363 GGGGACGATCAGTCCGCTTGGAGTCCAGTCTACTACTGAGCCGCC 412
   |||||
215 SGIAsnAspGluValArgPheGluLeuThrPheSerLeuAsnProG 232
   |||||
413 AGATAAAGTCAATGCTCCCTGGAGATGCCGATTCACAAACGGGTC 462
   |||||
232 IuLeuLysValIleAlaProtrPargGluTyrPdu.....Ile 244
   |||||
463 AAGGCGCGCAATGACCTGATGAGTACGCAAGCAACAGCGGATTC 512
   |||||
245 GluIleLysArgGluAspAlaIleGluTyrAlaLysLysLysIleAsnValProva 261
   |||||
513 CCCGCTACCTCCCAAGAACCCGTGAGCATGATGAGAACCTCATGCA 562
   |||||
261 IProValThrLysSerIleTyrSerIleArgAspIleAsnLeuTyrPdu 278
   |||||
563 TCAGCTACGAGGCTGCAATGCTGAGAGAACCCCAAGAACCGCTTCA 612
   |||||
278 euSerHisGluGlyAspLeuLeuGluAspProAlaAsnGluLysLys 294
   |||||
613 GGTCTCTTAC 621
   |||||
295 AspMetLys 297

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA53535

seq_documentation_block:

ID AA53535 standard; Protein: 146 AA.
 AC
 XX AA53535:
 XX

09-MAR-2001 (first entry)

Human colon cancer antigen protein sequence SEQ ID NO:1075.

Human: colon cancer; colon cancer antigen; diagnosis; detection;
 identification; cytosolic; cardioactive; neuroprotective; vulnary;
 immunomodulatory; muscular; gynaecological; gastrointestinal;
 nephrotropic; antibacterial; gene therapy; wound;
 neural disorder; immune system disorder; muscular disorder;
 reproductive disorder; gastrointestinal disorder; renal disorder;
 infectious disease; cardiovascular disorder.

Homo sapiens.

WO20005351-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05883.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI: 2000-587534/55.

N-PSDB; AAC98292.

Colon cancer associated gene sequences, referred to as colon cancer
 antigens, useful for the treatment, prevention, and diagnosis of colon
 disorders such as colon cancer -

Claim 11: Page 1660-1661; 2104pp; English.

AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 called human colon cancer antigens, given in AA53534 to AA54006. The
 human colon cancer antigens can have cytosolic, cardioactive, muscular;
 neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and
 vulnary; nephrotropic, anti-infective and antibacterial activities, and
 can be used in gene therapy. The colon cancer antigen polynucleotides,
 proteins and antibodies to the proteins are useful for the prevention,
 treatment and diagnosis of colon disorders, such as colon cancer. The
 polynucleotides may be used in diagnostics and research, such as for
 chromosome identification, and as hybridisation probes. The proteins
 may also be used to prevent diseases such as neural disorders, immune
 system disorders, muscular disorders, reproductive disorders,
 gastrointestinal disorders, wounds, renal disorders, infectious
 diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AA54007 represent sequences used in the exemplification of the present
 invention.

Sequence 146 AA:

alignment_scores: Quality: 463.00 Length: 99

Ratio: 4.926 Gaps: 0

Percent Similarity: 94.949 Percent Identity: 92.929

alignment_block:

US-09-775-693-1 x AA53535 ..

Align seg 1/1 to: AA53535 from: 1 to: 146

```

838 GGTATCTACGAGACCCAGCAGCAGCATCTTACCATGCTCATTTAGA 887
   |||
1 GlyThrSerGluThrProAlaGlyThrIleLeuTyrHisAlaHisLeuAs 17
   |||||
888 CATGAGGCTTTCACCATGAGACCGGAAAGTCCGCAAAATCAACACAGCC 937
   |||||
17 PileGluAlaPheThrMetAspArgGluValArgLysIleLysGluIle 34

```

938 TGGGCTGAATTTGCTGAGCTGCTGATACGCGTTAGCGCTACCT 987
 |||||||
 34 enGlyLeuLysPheAlaGlu***ValTyrThrGlyPheTrpHisSerPro 50
 |||||||
 988 GAGGTGAATTTGCTGAGCTGCTGATACGCGTTAGCGCTACCT 1037
 |||||||
 51 GlucylSerIlePheValAlaArgHisCysIleAlaLysSerGlnGluArgValG 67
 |||||||
 1038 AGGGAAGTGCAGGTGCTGCTGATACGCGTTAGCGCTACCT 1087
 |||||||
 67 uGlyLysValGlnValSerValLeuLysGlyGlnValTyrIleLeuGlyA 84
 |||||||
 1088 GGGAGTCCCGCTGCTGCTGATACGCGTTAGCGCTACCT 1134
 |||||||
 84 rGlySerProLeuSerLeuTyrAsnGlnGluLeuValSerMetLasp 99
 |||||||

seq_name: /SIDS2/gcycdata/geneseq/geneseqp/AA2001.DAT:AA16101

seq_documentation_block:
 ID AA16101 standard; Protein; 59 AA.
 AC AA16101;
 DE 12-OCT-2001 (first entry)
 XX Peptide #2535 encoded by probe for measuring cervical gene expression.
 XX Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX Homo sapiens.
 PN WO200157278-A2.
 PD 09-AUG-2001.
 PF 30-JAN-2001; 2001WO-US00670.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR.
 DR WPI; 2001-488901/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX Claim 27; SEQ ID No 20927; 487pp; English.
 XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs; see A110068-A128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 59 AA.

alignment_scores:

Quality: 228.00 Length: 55
 Ratio: 4.560 Gaps: 0
 Percent Similarity: 90.909 Percent Identity: 76.364
 alignment_block:
 US-09-775-693-1/rev x AA16101

Align seg 1/1 to: AA16101 from: 1 to: 59

573 CTCGAGCTGATGTCATGAGTTCTCATCCATGCTCCAGCGGCTTGG 524
 |||||||
 1 LeuIleAlaAspThrHisGlnValLeuValHisProAlaValLeuG 17
 |||||||
 523 GAGTACCGGGATGGGATCCCGGTTGCTTGGCTACTCCATCAGGTCA 474
 |||||||
 17 UserAspTrpAspGlyAsnProMetLeuLeuCyValPheHisGlnIle 34
 |||||||
 473 TTGGCGCCCTTGAACCGTTGTAGATTCAGGCATCCTCCAGGACCAT 424
 |||||||
 34 LeuIleAlaIleGluProValValGlnLeuGlnHisProGlySerAsn 50
 |||||||
 423 GACCTTATCTGGGG 409
 |||||||
 51 AspleuAsnLeuGly 55
 |||||||

seq_name: /SIDS2/gcycdata/geneseq/geneseqp/AA2001.DAT:AA28592

seq_documentation_block:
 ID AA28592 standard; Protein; 59 AA.
 AC AA28592;
 DE 17-OCT-2001 (first entry)
 XX Peptide #2629 encoded by probe for measuring placental gene expression.
 XX Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 XX Homo sapiens.
 PN WO200157272-A2.
 PD 09-AUG-2001.
 PF 30-JAN-2001; 2001WO-US00663.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR.
 DR WPI; 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX Claim 27; SEQ ID No 28861; 654pp; English.
 XX The present invention relates to single exon nucleic acid probes (SENPs;
 CC see A113115-A1157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.

(see AAI00010-AAI10067). The present sequence is a peptide encoding 27 amino acids. The probes are useful for measuring human gene expression in

XX

1050 roArgThr.....CysArgAlaThrSerSerAlaSer 1061
 1073 TGTACATCTCGCCGAGATCCCACTGTC.....TCTAC 1110
 ||| |||||
 1062 Cys.....ArgCysLeuSerCysProGlnSerTh 1071
 1111 AATGAGAGCTGG 1123
 ||| |||||
 1071 rTrnAlaAlaTrp 1075

seq_name: /stds2/gcdata/geneseq/geneseqp/AA2001.DAT:AA59813

seq_documentation_block:
 ID AAB59813 standard; Protein; 1017 AA.

XX AAB59813;
 AC
 DT 04-APR-2001 (first entry)
 XX
 DE Tuto protein #4.
 XX
 KW Toluene degradation; enzyme; waste degradation; Tuto.
 XX
 OS Thauera aromatica.
 OS Xanthomonas maltophilia.
 OS Geobacter metallireducens.
 OS Azarcus toluelyticus.
 XX
 PN W0200072650-A2.
 XX
 PD 07-DEC-2000.
 PF 24-MAY-2000; 2000WO-US14298.
 XX
 PR 01-JUN-1999; 99US-0323872.
 XX
 PA (UYOH-) UNIV OHIO.
 XX
 PI Coschigano PW;
 DR WPI: 2001-041080/05.
 DR N-PSDB: AAF23625, AAF23627.
 XX
 PT Composition comprising toluene degrading enzyme useful for biological
 PT treatment of organic compounds, especially for degrading toluene or its
 PT analogs
 XX
 PS Disclosure; Fig 5; 122pp; English.
 XX
 CC The present invention relates to toluene degrading enzyme genes and
 CC proteins tuth (see AAF23629 and AAB59831), tuti (AAF23630 and AAB59832),
 CC tutf (AAF23631 and AAB59833) and tutg (AAF23632 and AAB59834). The
 CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
 CC toluene degrading enzymes are useful for biological treatment of organic
 CC compounds and in particular for the degradation of toluene and its
 CC analogs contained in liquid or solid waste source. The present sequence
 CC is a protein sequence for toluene degrading enzyme, Tuto.
 XX
 SQ Sequence 1017 AA;

alignment_scores:
 Quality: 193.00 Length: 497
 Ratio: 0.910 Gaps: 29
 Percent Similarity: 42.656 Percent Identity: 23.944

alignment_block:
 US-09-775-693-1 x AAB59813 ..

Align seg 1/1 to: AAB59813 from: 1 to: 1017

38 GCGCCTGGACACCTCTGTCATCTCTGCTGCTGCTGAAGACACAGCTAT 87
 :||| ||||| :|||

4 SerAlaThrThrProLysIleProThrCys.....SerArgCysI 17
 88 GACGTCATTGCTCTATCTGCGCA.....ACATTGGCCAGAAAGACTT 131
 : ||| ||||| ||| :|||
 17 eProAncysProThrTrpProCysArgThrThrCysGlnAlaThrThra 34
 132 CGAGG...AAGCCAGAGAGAGAGCAGCTGAGCTTGGGCCA...AAAAG 175
 ||| :||| ||||| :||| :|||
 34 rArgSerArgProThrArgArgArgSerMetAsnThrGlySerArg 50
 176 TGTTCATTGAGATGTCAGCAGGAGCTTG..... 205
 :||| ||| :||| :|||
 51 IleAlaCysArgAlaSerValSerProIleSerIleArgGlnThrSerAl 67
 206TGAGAGATTTCATCTGGCCGCCCTCA 233
 :||| ||||| :||| :|||
 67 AlaAlaCysIleArgSerAlaAlaTrpArgArgProSerLeuProValTr 84
 234 GTCCAGCGCCTGTTATGAGAGCCGCTACCTCTG..... 268
 :||| ||| ||| |||||
 84 hrThrAlaSerCysArgProThrLysProSerThrLysGlyCysTrp 100
 269 ...GCACCTCTCTGCGCAGCGCTCATCGCC..... 298
 :||| ||| :||| :|||
 101 ArgAlaSerSerSerProLysSerIleSerProLysProArgProThrCy 117
 298 298
 117 sarArgProSerProGlyThrAlaArgArgValSerThrThrSerProArg 134
 299GCANACAAGTGAATGCCCCAGCGGAG..... 328
 :||| :||| :||| :||| :|||
 134 eThrThrGlyArgArgTrpSerSerProAlaArgArgSerAlaGlyArg 150
 329 ...GGGCCAAGTATGTGTCGCCAGCGGCCACAGAAAGGGA...ACGAT 372
 ||| :||| ||| :||| :||| :|||
 151 AlaGlyArgAlaGlyCysAlaArgSerSerArgLysThrSerArgProI 167
 373 CAGTCCGGTTGAGCTCA.....GCTGCTACTACT 404
 :||| :||| :||| :||| :|||
 167 eArgSerAlaArgProSerCysSerLysSerProThrSerValSerAlaP 184
 405 GGCCCCCAGCA.....TAAAGTCATTG 427
 :||| :||| :||| :||| :|||
 184 heProProSerProAlaArgAlaSerArgThrArgCysArgArgAsnSer 200
 428 CTCCTGGAGGA..... 439
 :||| :||| :||| :||| :|||
 201 LeuProSerSerValThrArgSerSerAlaThrArgAlaAlaThrProAr 217
 440TGCTGAATTCTACACC..... 457
 :||| :||| :||| :||| :|||
 217 gArgLysThrProCysGlyArgThrThrArgProProSerSerThrAla 234
 458GPTCAAGGCGCCAGTACCTGANG...AGTACCAAGACAC 499
 :||| :||| :||| :||| :||| :|||
 234 rGAsnSerSerArgAlaThrTrpMetLArgTrpAsnSerSerArgTrpAsn 250
 500 ACGGATTCCATCCCGGTCA.....CTCCAGAAACCCGTG 537
 :||| :||| :||| :||| :||| :|||
 251 ValArgPheProSerMetAlaProAlaSerArgAlaProThrAlaLysSe 267
 538 AGCATGATGAGAACCTCATGCACA.....TCAGTACGAGGCTGG 578
 :||| :||| :||| :||| :||| :|||
 267 rSerArgLysArgThrIleCysSerSerSerProSerAlaAlaProThrP 284
 579 AATCTTGAGAACCCCAAGAACCAAGCCCTCCAGGTCTTCACACGAGA 628
 :||| :||| :||| :||| :||| :|||
 284 roArgAlaArgThrProAlaThrThrProThrPro...SerSerArgGln 299
 629 CCC.....AGGACCGAGCCCAAGCCCAACACCCCTGACA..... 664
 :||| :||| :||| :||| :||| :|||
 300 ProSerGlySerAlaArgProSerProProSerSerSerAlaIleProAr 316

[illegible]

```

398 .....ACT 400
641 eRLaGlyCysArgAlaSerProAlaHisArgGlyArgProAlaArgThr 657
401 CACTGCCCCCAGATTAAGTTCATTGCTCCGAGAGATGCCGTAATTC 450
568 SerArgProAlaArgProArgProAlaCysSerArgPhcysArgGlnPr 674
451 TACACCGGTTCAAGGCCCATGACTGATGAGT .....ACGC 491
674 calThrtPrArgArgAlaAlaAlaThrProThrAlaProLeuAlaPr 691
492 AAGCAGACGGGATTC.....CCATCCGGTCACTCCCA 526
691 roSerArgGlyGlyValAlaAlaValAlaArgProHisArgProProAla 707
527 AGAACCCCTGGAGCATGATGAGACCTCATCAGATCAGTACGAGGT 576
708 ArgProArgGlyLeuAlaProAlaHisAlaCysProProProThrArgPr 724
577 GGAATCTGTGAGAACCCCAAG.....ACCA 602
724 oGlnSer...ArgAlaProArgAlaGlyAspArgGlnThrGlyValAla 740
603 AGGCCTCCAGGTCCTCTACAGACGACCCAGACCCCAAG..... 646
740 rGArgLeuCysArgGlyLeuArgSerProArgAlaAlaProArgArgArg 756
647 .....CCCCAACCCCTGACATTCGAGATCGAGTCAAA 684
757 GlyArgProCysProThrSerProGlyCysArgAlaAlaGlnProGlyArg 773
685 AAGGGGCTCTGAGAGTGACCAACGTCACGATGGACCCACCAC 733
773 gArgGly.....GlyArgCysGlyProGlyThr 783
734 .....AGACCTCTGAGACCTTCATGT..... 757
783 laGlyGlyThrSerArgProPro...SerGlyProCysArgProArgAla 798
758 .....ACCTG 762
799 ValThrThrAlaProPhLeuGlnProThrAspProAlaAlaProSerSe 815
753 AACGAGTCCGGGACAGCATGGCGTG.....GCCGTAT 797
815 ArgSerSerArgSerProArgSerThrArgThrCysArgCysSerVala 832
798 TGACATCTGTGAGAACCGCTTCATGTGAGTGAAGTCCGAGGTATCTACG 847
832 rGSerSerThrPro.....GlyGlyArgProCysThr 842
848 AGACCCCGACAGCAGCATCTTACATCTCATTTACATTCAGAGGCC 897
843 ArgPro.....GlyProGlyAlaArgAlaArgAlaThrLeuProCysPr 857
898 TTCACATGAGCCGGAAGTGGCAAAATCAACAGAGCCCTGGGCTTGA 947
857 calaProThrProArg..... 862
948 ATTGCTGACGTGTGTATACGGGTTTACGGCTAGCCGTAGTGAAT 997
863 .....ProSerLeuGlyProAla...ArgCysPro 871
998 TTGTCCGCACTGCATGCCCAAGT.....CCAG 1026
872 LeuGlyAlaProAlaProAlaProAlaProAlaProAlaThrArgProAl 888
1027 GAGCGAGTGAAGGAAAGTCAAGGCTCCGCTCAGGCGGCTGTA 1076
888 aglyAlaArgArgArgArgSerArgCysAla..... 898
1077 CATCTCGGCGGAGTCCCACTGTCTCTACATTAAGAGAGCTGTGA 1126

```

```

899 .....CysArgSerThrGlyArg..... 904
1127 GCATGACGTGCGAGGTGATTAAGCCAACTGATGCGACCGGTTTCATC 1176
905 .....ProAlaArgArgAlaSerArgGlnGlyProProGlySerThr 919
1177 AACATCA 1183
919 RasPser 921

```

```

seq_name: /SID2/gcdata/geneseq/geneseq/AA1999.DAT:AAW87503
seq_documentation_block:
ID AAW87503 standard; Protein; 1212 AA.

```

```

AAW87503:
23-FEB-1999 (first entry)
Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.
KW Human; N-methyl-D-aspartate receptor; NMDAR2C;
OS Homo sapiens.
PN US5849895-A.
PD 15-DEC-1998.
PR 20-APR-1994; 94US-0231193.
PR 20-APR-1994; 94US-0231193.
PR 20-APR-1993; 93US-0052449.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Daggett LP, Lu C;
PI WPI; 1999-069812/06.
DR N-PDB; AAW82909.
XX DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
XX assembly of functional glutamate receptor subunits
XX
XX Example 3; Columns 253-262; 203pp; English.
XX
XX The present sequence represents a human N-methyl-D-aspartate (NMDA)
XX receptor subunit (NMDAR). The nucleic acid sequence does not contain
XX the 366 5'-most nucleotides, by the insertion of 11 nucleotides between
XX nucleotides 1300 and 1301, nor the 15 nucleotides at positions
XX 1960-1974, nor the 1061 3' nucleotides, as set forth in AAW82889. The
XX CDNA sequence is derived from clone NMDA21. The NMDAR subunits contribute
XX to the formation of NMDA-activated cation-selective ion channels. In
XX addition to being useful for the production of NMDA receptor subunit
XX proteins, the nucleic acids are also useful as probes to identify and
XX isolate nucleic acids encoding related receptor subunits. Functional
XX glutamate receptors can be assembled from several NMDA receptor subunit
XX proteins of one type (homomeric) or from combinations of subunit proteins
XX of different types (heteromeric). The present invention also comprises
XX methods for using such receptor subunits to identify and characterise
XX compounds which affect the function of such receptors, e.g. agonists,
XX antagonists and modulators of glutamate receptor function. The invention
XX also comprises methods for determining whether unknown protein(s) are
XX functional as NMDA receptor subunits.
XX
XX Sequence 1212 AA:

```

```

alignment_scores:      Length: 469
Quality: 174.00      Gaps: 25
Ratio: 0.961

```

AAW8/503

FROM: 1 to: 1212

38 GCGGCTGGAGACACT.....CG 54
671 AlaSerIleProLeuLysAlaArgSerLeuLeuProLeuAlaIle 687
55 TGCAATCTGTGTGGTGAAGAACAAGGCTATGACGTCAATTGCCATTCT 104
687 TAlaSerProCys.....ArgAlaThrProThrLeuSerGlyProThr 702
105 GGCCAACTATGGCCAGAAAGAACCTTCGAG...AASCCAGAAAGAG 151
702 IPrArgSerCysSerIleTrpLysAlaHisArgAsnTrpAlaGln 718
152 CACTGAAGCTTGGGCCAAAAGGTGTGCATGGAGATGTCACAGGGAG 201
719 CysGlySerGlnGlySerAlaArg.....MetArgIleThrArgAlaIle 733
202 TTGTGGAGAGTTCAATCGGGCGGCATCCAGTCCAGGCGCACTGNATGA 251
733 aserTrpHisSerThrThrIleAlaIleSer..... 744
252 GGAGCGTAGCTCTGGCACCTCTTGTGCACAGCCCTCATCGCCGCA 301
745 ..ThrcyStrIprTrpProThrPoliyrProCysTrpSerSerProGly 360
302 AACAAAGGAAATTCGCCAGCGGGAGGGGCCAAATATGTGTCCACGCG 351
761 SerThrTrpSer.....ThrglysercysAlaThrAlaCysCysProThrH 775
352 GCCACAGAAAGGGGAGC...ATCAGTCTGGTTTACGTCAGCTGCT.. 397
775 sProSerTrpHisSerCysStrpLeuSerAlaGlyAlaSerThrAlaAlaIS 792
398 792
792 eAlaGlyCysAlaGlaSerProAlaHisAlaGlyIleArgProAlaIle 808
401 CACTGGCCCCCAGATMAAGTCATGTGCTCCGTGGAGATGCGTAATTC 859
809 SerArgProAlaIleArgProArgProAlaCysSerArgIlePhcysArgGlnPr 825
451 TACAACCGGATCAAGGGGCCGATAGCTGATGAGAT.....ACGC 491
825 oAlaThrTrpProAlaGlaAlaIleAlaThrProThrPhrAlaProLeuAla 842
492 AAAGCAACGGGATTC.....CCATCCCGGTCACTCCCA 842
842 roserArgIleIleValAlaAlaAlaValArgProHisArgProProAla 858
527 ACAACCCTGGAGCATGATGAGAACCATCATGACACATACAGAGGT 576
859 ArgProArgGlyLeuAlaProAlaHis..... 646
577 GGAATCTCGGAGAACCCCAAG.....ACCA 602
875 oGlInsEr..ArgAlaProArgGlaGlyAspArgGlnThrGlyValAla 891
603 AGGCCCTCCAGGCTCTCTACAGAGAGACCCGAGCCACGCAAG..... 646
891 rgaIleuCysAlaGlyLeuArgSerProAlaIleAlaIleProArgArg 907
647CCCCAAACCCCTGACATTTCTGAGATCGAGTTCAAA 684
908 GlyArgProCysProThrSerProLeuLucysArgAlaAlaInProLys 924
685 AAAGGGGTCCCTGGAAGTGACCAACCAAGATGGACCAACCCAC 733

```

924 gArgLy.....GlyArgGslProGlyThra 934
724 .....AGACTCTTGAGAGCTCTCATGT
934 laGlyLyThSerArgProPro...SerGlyProCysArgProAla 949
758 .....
950 ValThrThraAlaProPheLeuGluProThrAspProAlaAlaProSerSe 962
763 AACGAAGTCGGGGGCAAGCATGCCGTG.....GCCGTAT 797
966 rArgSerSerThrSerProArgSerThrPatGlyThCysArgCysSerVala 983
798 tGACATGCTGAGAACACCGCTCATTTGAAATGAAGCCCGAGATATCAGC 847
983 rGseSerThrProPro.....GlyGlyArgProCysThr 993
848 AGACCCGAGAGGACACATCTCTTACCATGCTCATTTAGACATCGAGGCG 897
994 ArgPro.....GlyProGlyAlaArgAlaArgValThrLeuProCysPr 1008
898 TTACCATGTCGACCGGGAAGTCGCAAAATCAACAACAGCGCTGGCTTAA 947
1008 oAlaProThrProArg.....
948 ATTTCGTAGCTGTGTATACCGGTTCATGCGGCTACGCTGAGTGTGAT 1013
1014 .....ProSerLeuGlyProAla...ArgCysPro 1022
998 TTGTCCGACCATGTCATCGCCAAGT.....CCGAG 1026
1023 LeuGlyAlaProAlaProAlaProAlaProAlaThrAlaThrArgProAl 1039
1027 GAGGAGATGTGAAGGAAGACTGCAAGTGTCCGCTCTCAAGGCAAGGTGTA 1076
1039 aGlyAlaThrPatArgArgSerArgCysAla..... 1049
1077 CACTCTCGCGCGGAGTCCCATCTCTCTCTACAAATGAGAGCTGTGTA 1126
1050 .....CysArgSerThrGlyArg..... 1055
1127 GCATGACGTGACAGGTGATTATGAGCCACTGATGCGACGGGTTCATC 1176
1056 .....ProAlaArgAlaArgAlaSerArgGlyProProGlySerThr 1070
1177 AACATCA 1183
1070 rAspSer 1072

seq_name: /SIDS2/gcgcdata/geneseq/geneseq/Aa2000.DAT:AA26240
seq_documentation_block:
ID AAB26240 standard; Protein: 1081 AA.
AC AAB26240;
XX
XX
DT 23-FEB-2001 (first entry)
DE
XX Human N-methyl-D-aspartate receptor subunit NMDAR1A #18.
XX
XX Human; N-methyl-D-aspartate receptor; NMDA, NMDAR1A; Ionotropic;
KW glutamate receptor; drug screening; animal model; disease diagnosis;
KM genetic screening.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 116 /label= unknown
FT /note= encoded by TGA*
FT Misc-difference 140 /label= unknown

```


FT	Misc-difference	/note= "encoded by TGA"
FT	162	
FT	/label= unknown	
FT	/note= "encoded by TGA"	
FT	183	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGA"	
FT	258	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGA"	
FT	274	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGA"	
FT	300	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TAG"	
FT	313	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGA"	
FT	337..338	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGATGA"	
FT	431	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGA"	
FT	561	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TAG"	
FT	593..594	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGATGA"	
FT	692	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGA"	
FT	697	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TAA"	
FT	858	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGA"	
FT	1001	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TAG"	
FT	1058	
FT	Misc-difference	/label= unknown
FT	/note= "encoded by TGA"	
FT	1076	
FT	Misc-difference	/label= UNKNOWN
FT	/note= "encoded by TGA"	
FT	/note= "the coding sequence of this protein contains a number of in-frame stop codons which are represented by Xaa in the protein sequence"	
PN	US6111091-A.	
XX		
PD	29-AUG-2000.	
XX		
PF	29-SEP-1997;	97US-0940086.
XX		
PR	20-APR-1994;	94US-0231193.
PR	20-APR-1993;	93US-0052449.
XX		
PA	(MERI) MERCK & CO INC.	
XX		
PI	Daggett LP, Lu C;	
XX		
DR	WPI; 2000-578607/54.	
DR	N-PSDB; AAA95033.	
PT	Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit	
PT	for identifying mutations and for developing drugs against various	
PT	disease states -	
XX		

[illegible]

745 ProArgAlaGlyAspArgInThrGlyValAlaArgArgLeuCysAlaG 761
619 TACACGAGACCCAGACCCAGCCAAAG.....CCCC 650
: ||| |||||: |||: |||
761 yLeuArgSerProArgAlaAlaProArgArgGlyArgProCysPro 778
651 CAACACCCCTGACATTCGAGATCGATTCAAAAAAGGGCTCCGTGA 700
: ||| : : : : : |||
778 hrSerProGlyCysArgAlaAlaGlnProGlyArgGly..... 791
701 AGGTGACCAACGTCAAGATGCGACCCACC.....AG 735
792GlyArgCysGlyProGlyThrAlaGlyGlyThrSerAr 804
736 ACCTCCTTGAGGCTTCATGT..... 757
||| |||: |||
804 gProPro..SerGlyProCysArgProArgAlaValInThrAlaPro 820
758ACCTGAACGAAGTCGCGGCA 778
820 heLeuGlnProThrAspProAlaAlaProSerSerArgSerSerArg 836
779 AGCATGGCGCTG.....GCCGATTGACATCGTGAGAAC 813
: ||| : ||| : ||| |||
837 ProArgSerThrParGThrCysArgCysSerValArgSerSerThrPro 853
814 CGCTTCATTGGAATGAAGTCCGAGTATCTACGAGACCCAGCAGCAC 863
: : : : ||||| |||
853 yGlyArg.....ProCys**ThrArgPro.....GlyP 863
864 CATCCTTTACCATGCTCATTTAGACATCGAGGCTTACCATTGACCGGG 913
: : : : ||||| |||
863 roGlyAlaArgAlaArgValThrLeuProCysProAlaProTrpProArg 879
914 AAGTGGCAAAATCAACAAGAAGCGCTGGCTGAAATTTGCTGAGCTGTG 963
879 879
964 TATACGGGTTTACGCCCTAGACCTGAGTGAATTTGCCGCACATGCAT 1013
||| ||| |||: ||| |||
880ProSerLeuGlyProAla..ArgCysProLeuGlyAlaProAlaP 894
1014 CGCAGGT.....CCGAGGAGCGAGTGAAGGA 1042
|||: : : : ||| : : : |||
894 roProAlaProAlaProThrAlaThrArgProAlaGlyAlaTrpArgArg 910
1043 AAGTGCAGGTGCCCTCTCAAGGGCCAGGTACTCTCGCGCGGAG 1092
: : : |||||
911 ArgSerArgCysAla..... 915
1093 TCCCACTGTCTCTTACAATGAGAGCTGTGAGCATGACGTGACGGG 1142
||| ||||| ||| : |||
916CysArgSerThrGlyArg.....ProAlaArgA 925
1143 TGATTATGAGCACTGATGCCACCGGGTTTCATCAACATCA 1183
||| ||||| |||: |||
925 rGAlaSerArgGlnGlyProProGlySerThrAspSer 938


```

192 .....:|||||
469 CGCAATGACCTGATGAGTACGCAAGCAACAGCGGATTCCTCCGGT 518
195 rolaaglyproglyly.....SerArgaspProglyProPro 207
519 CACTCCCAAGAACCCGTGAGCATGATGAGAACCTCATGACATCAGCT 568
208 gLyAlaInglyProAlaInglyProgly..... 216
569 ACAGGCTGGAATCTCGAGAA.....CCCCAAGAACCA 603
217 .glySerArgaspProglyProProglyAlaInglyProAlaInglyProG 233
604 GCGCTCCAGGCTCTACAGAACAGCCAGGACCC.....AGCCAAAGC 647
233 LyglySerArg.....AspProglyProProglyAlaIngly 245
648 CCCCACACCCCTGACATCTCGAGATGAGTTCAAAAAAGGCGTCCGTG 697
246 ProAlaInglyProglylySerArgasp.....ProglyPro... 257
698 TGAAGTGAACCAAGTCACAGATGACACCCAGACCTCTTGAG 747
258 .....ProglyAlaHisglyProAlaInglyProlysglyA 269
748 CTCTTCATGATCTGACAGAGTCCGCGGCAAGCATGCGCGCTGAT 797
269 la...HisglyProAlaInglyProlysglyAla.....HisglyProAla 282
798 TGACATCTGAGAGAACCCCTTATGGAATGAGTCCGAGGTACTGACG 847
283 GlyProlysgly.....AlaHisglyProAlaInglyProlysglyAla 297
848 AGACCCCGCAGGACCATCTTTACATGCTATTGACATGACAGGC 897
297 nglyProAla..... 300
898 TTCACCATGACCCGGAAGTCCGCAAAATCAAAAGCGCTGAGTGA 947
301 .....glyProglylySerArgaspProglyProProglyAlaIng 314
948 ATTCTGAGCTGCTGATACCGCTTTACGCGCTAGCCCTGAGTGAAT 997
315 ..glyProAlaIngly..... 318
998 TTGTCCGCGCATGCGCAAGTCCGAGAGCGATGAGAGCAATG 1047
319 .....ProglylySerArgaspProglyProProglyAlaIngly 332
1048 CAGGTGTCGCTCAGAGGCGAGGTATCTCGCGCGGAGTCC 1097
333 .....ProAlaInglyProglylySerArgaspProglyProPro 345

```

seq_name: /cgn2_6/ptodata/2/1aa/bb_comb.pep:US-09-219-849-6

seq_documentation_block:

Sequence 6, Application US/09219849

Patent No. 6150081

GENERAL INFORMATION:

APPLICANT: VAN HEERDE, GEORGE V.

APPLICANT: VAN RIJN, ALEXIS C.

APPLICANT: BOUWSTRA, JAN B.

APPLICANT: DE WOLF, FREDERIK A.

APPLICANT: MOEBROEK, ANDREAS

APPLICANT: WERTEN, MARC W.T.

APPLICANT: WIND, RICHELIE D.

APPLICANT: VAN DEN BOSCH, TANJA J.

TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN

TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE

PREPARATION THEREOF

FILE REFERENCE: 2728-2

```

CURRENT APPLICATION NUMBER: US/09/219,849
CURRENT FILING DATE: 1998-12-23
NUMBER OF SEQ. ID NOS.: 50
SOFTWARE: Patentn Ver. 2.1
SEQ ID NO: 6
LENGTH: 960
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Illustrative
US-09-219-849-6

```

alignment_scores:

Quality	139.50	Length	357
Ratio	1.065	Gaps	20
Percent Similarity	36.695	Percent Identity	26.050

alignment_block:

US-09-775-693-1 x US-09-219-849-6 ..

Align seg 1/1 to: US-09-219-849-6 from: 1 to: 960

```

131 TCAGAGAACCCGGA.....AGAGGCACTGAGCTTGGGCCCAAAAG 174
||||| |||||
216 SerArgaspProglyProProglyAlaHisglyProAlaInglyProlysgl 232
175 GTGTCATGAGATGTCACGAGGAGCTTTGGAGGAGTTCATGCGCC 224
232 Y.....AlaHisglyProAlaInglyProlysglyAlaHisg 244
225 GGCATCCAGTCA.....GGCAGCTGATGAGAGACCGCTACTCTCG 268
||||| |||||
244 LyProAlaInglyProlysglyAlaHis..... 252
269 GCACCTCTTGGCAGGCGCTGATCGCCCGCAACAGTGAATGCC 318
||||| |||||
319 CAGGGGAGGGGCCAAGTATGTCCTCCAGCGGCCACAGAAAGGGA 368
||||| |||||
263 oAlaInglyProglylySerArgaspProglyProProglyAlaIngly 280
369 CGATCAGTCCGGTTGAGCTGAGCTGCTACTGCGCCGCCCAATAA 418
280 ro. AlaInglyProglylySerArg.....AspProglyProProgly.. 293
419 AGTCAATGCTCCCTGAGAGATGCTGAATTTCAACACGCTTCAAGGC 468
294 .....AlaInglyP 297
469 CGCAATGACCTGATGAGTACGCAA.....GCACACGG 503
297 rolaaglyProglylySerArgaspProglyProProglyAlaIngly 313
504 GATTCACATCCGCTCAGTCCCAAGACCGTGGACATGAGTGAAGC 553
314 ProAlaInglyProglylySerArgaspProglyProProglyAlaIng 330
554 TCATGACATCAGTACGAGGCTGAATCTGAGAA..... 590
330 yProAlaInglyProglylySerArgaspProglyProProglyAlaIng 347
591 ..CCCAAGAACAGCGCTCAGTCTCTACAGCAACAGGAGGAGCC 638
347 LyProAlaInglyProglylySerArg.....AspProglyPro 359
639 .....AGCCAAAGCCCAACACCCCTGACATTTCTGAGATGAGTCA 682
360 ProglyAlaInglyProAlaInglyProglylySerArgasp..... 373
683 AAAAGGGTCCC.....TGTAAGTGACCAACGTCAGAGATGACGC 726

```

557 TGCACAT CAGCTACGAGGCTGGA :
||||||| ::||| ||| ::

```

521 saiahiProtyrAlaMetSerProSerLeuGlySerLeuArgProTyr 538
580 ATCTGGAGAACCCCAA...GAACCAAGGCGCTCCAGGCTCTTAACGAA 626
538 roProGlyProAlaHisLeuProProHisSerGlnValSerTyrSer 554
627 GACCCAGAGACCCCAAGCCCA...ACACCCCTGACATTCCTCGAA 673
555 GlnAlaGlyProAsn..GlyProProValSerSerSerSerSerSer 570
674 TCAGATTCAAAAAGGGTCCCTGTGAAGGTGACCAAGCTAAGGAGGC 723
571 SerSerThrSerGlnSerTyr.....ProCysSerHisProSer 584
724 ACCACCCACCAAGCT.....CCTGGAGCTCTTCATGTA 758
584 rProSerGlnGlyProGlnGlyAlaProTyrProPheProProValPro 601
759 CCTGAACGAGTCGGCGGCAAGCATGGCGTGGCCGTATTGACA..... 802
601 hValThrThrSerSerAlaThrLeuSerThrValIleAlaThrValAla 617
803 .....TCGTGGAGAACCGCTTCATTGGAATGAATGATCCCGAGT 840
618 SerSerProAlaGlyTyrTyrThrAlaSer.....ProCysSerHisPro 627
841 ATCTACGAGACCCCAAGCAAGCATCTTACCATCTCATTTAGACAT 890
628 .....ProProGlyProProProTyr.....G 635
891 CGAGGCTTCACATGACCGCGGAGTGGCAAAATCAACAGAGCCCTGG 940
635 LysLysArgAlaPro.SerProGlyAlaTyrTyrThrAlaThrProProG 651
941 CTTGAATTTGCTGACGTGGTGTATACCGGTTACGCGCTACCGCTGAG 990
651 YTYrLys.....ProGly..... 655
991 TGTGAATTTGCCGCGCATGCTGCATCGC.....CAAGTCCAGAGGAGCT 1034
656 .....SerProProSerPheArgThrGlyThrProProGlyTyrArg 669
1035 GGAAGGCAAGTGCAGGTGTCGCTCAAGGCGCAGTGTACATCCCTGG 1084
670 GlyThrSerPro.....ProAlaGlyProGlyThrPhe...Ly 681
1085 GCCGGAGTGCACGCTGCTCTCTTACATGAGGAGCTGTGAGCATGAAC 1134
681 sProGlySerProThrVal.....GlyProGlyProLeuProP 694
1135 GTGCAGGG 1142
694 roAlaGly 696

```

seq_name: /cgn2_6/plodata/2/1aa/5B.COMB.pep:US-08-660-963-12

seq documentation block:
 : Sequence 12, Application US/08660963
 : Patent No. 5852187
 : GENERAL INFORMATION:
 : APPLICANT: Thorne, Michael O.
 : APPLICANT: Gaylin, Bruce D.
 : APPLICANT: Horikawa, Reiko
 : APPLICANT: Lyons Jr., Charles E.
 : TITLE OF INVENTION: MOLECULAR CLONING OF THE OVINE PITUITARY
 : NUMBER OF INVENTION: GROWTH HORMONE RELEASING HORMONE RECEPTOR
 : CORRESPONDENCE ADDRESS: 18
 : ADDRESSEE: POPHAM, HAIK, SCHNOBICH & KAUFMAN, LTD.
 : STREET: Metropolitan Square Building, Suite 800, 1450
 : CITY: Washington
 : STATE: D.C.

```

? COUNTRY: USA
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/660,963
? FILING DATE: 12-JUN-1996
? CLASSIFICATION: 536
? ATTORNEY/AGENT INFORMATION:
? NAME: O'Shaughnessy, Brian P.
? REGISTRATION NUMBER: 32,747
? REFERENCE/DOCKET NUMBER: 18046.036
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-824-8000
? TELEFAX: 202-824-8199
? TELEX: 248516
? INFORMATION FOR SEQ ID NO: 12:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 498 amino acids
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-660-963-12

alignment_scores:
  Quality: 137.50      Length: 437
  Ratio: 0.795         Gaps: 25
  Percent Similarity: 39.588      Percent Identity: 22.197

alignment_block:
  US-09-775-693-1 x US-08-660-963-12 ..

Align seq 1/1 to: US-08-660-963-12 from: 1 to: 498

128 ACTTCGAGAGACCCAGAGAGAGGCACTGAGCTTGGGCCCAAAAGGTG 177
   |||:.. ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  27 ThrThrGlyAspGlyGlyArg.....SerThrGlyProProArgLe 40
   :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :..
  178 TTCATTGAGAGATGTCAGAGGGAGTTG.....TGAAGAGTTGATCTG 221
   :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :..
  40 uThr...GlnAlaAlaValAlaGlnLeuProProThrAlaAlaGlyCys 56
   :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :..
  222 GCCGGCATTCAGTCCAGCGCATGTATGAGAGCCGCTACCTCGTGACA 271
   || |||:.. ||| ||| ||| ||| ||| ||| ||| ||| |||
  56 LysAlaProAlaSerSerAlaCysTrpAlaProCysGlnSerSerTrpAla 72
  271 .....
  73 ThrCysThrGlnSerValMetSerSerLeuSerGlnuArgThrSerTyr 89
  272 .....CCTCTTGGCAAGGCCCTGCATCGCCGCAAGCAAGTGAAT 314
   :.. ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  89 sValTyrLysLeuLeuLysGly.....CysProThrProProThrAla 104
  315 GCCCCAGGAGGGGGCCAGAGTATGTGCCACAGGGGCCACAG..... 358
   :.. ||| ||| ||| :.. :.. ||| ||| ||| ||| ||| |||
  104 LPro..GlySerGlyThrGlyCysAlaGlyArgTrpGlnAlaLeu 119
  359 GAAAGGGAGAGAGTCCGCGTTTGAGCTGACGCTGCTACCTACGAGCC 408
   || |||:.. :.. ||| ||| ||| ||| ||| ||| ||| |||
  120 GluSerGlyAlaSerProAlaArgLeuSerSerLeuThrSerAlaTrp 136
  409 C.....
  136 rGlnGlyLeuArgGlyThrAlaProLeuGlnAlaGlyArgSerProSer 153
  410 .....CCGAGTAAAGTCAATGCTCCCTGAGAGAGCCG...AAT 448
   ||||| ||||| :.. ||| ||| :.. ||||| |||

```

```

153 rgleuileproarg...Prohaleucyprotrpsercysleuargarg 168
449 TCTACACCGGTTCAAGGCGGCAATGACCTGATGAGTACGCAAGC... 496
169 AsnProThrSerProArgIlySerSerThrProThrAlaThrAlaSerAr 185
497 ...AACACGGATTCACATCCCGGTCACTCCCAAGAACCGGTGAGCAT 542
185 gleuGlnProSerThrProSerSerSerThrProSerLeuSerIly... 200
543 GATGAGAACCTCATGACATCAGCTACGAGGCTGGAATCCTGAGAAC 592
201 ...SerThrAlaProGlyThrThrSerThrProSerCysSerPro 214
593 CCAAGAACCAAGCGCTCCAGGCTCTACAGAGAACG... 631
215 ProLeuSerSerArgArgGlnLeuGlySerArgThrProProSerPheTh 231
632 ...AGGACCCAGCCCAAGCCCAAGCCCGCTGACATTCGAGATGCA 677
231 rGlyArgThrThrThrThrAlaAlaSerProLeuSer... 243
678 GTTCAAAAAAGGGTCCCTGTGAGAGTGACCAACGTCAGATGAGACCA 727
244 ...CysAlaArgLeuLeuPro 249
728 CCC...ACCAGACTCTCT...TGGAGCTCTTCATG 756
250 ProLeuLeuSerArgProProThrSerAlaGlyCysThrGlnLeuLeu 266
757 TACCTGAACGAG... 769
266 sThrProAlaSerProProHisCysProAlaGlnGlySerSerIlyG 283
770 ...TCGCGGCGACACATGCGC 787
283 LYTTPheserLeuProGlyGlyPheLeuGlySerSerSerProAlaCysGly 299
788 TGGCCCGATTTGACATGCTGAGAACCGCTTATGATGATGATGCCGA 837
300 Trp...ValAlaSerTrpProLeuLysMetLeuArgAlaGlyThrTr 314
838 GGTATCTAGAGAGCC...CACGAGGACACATCC... 868
314 pThrThrAlaProProThrGlyGlySerSerLysAspProSerSerI 331
868 ... 868
331 euleuGlyThrLeuGlyPhePheSerIleLeuSerValSerCysGlyAsn 347
869 TTTACCATGCTCATTTAGACATCGAGCGCTTCACCATGAGCGGG... 913
348 TrpSerGlnLeuArgAlaAlaSerThrProSerLeuSerThrGlyValSe 364
914 ...AAGTGGCAAAATCAAAACAGGCT 938
364 rLeuSerGlnArgPheSerSerPheArgCysLeuGlnSerThrMetSerP 381
939 GGGCTTGAATTTGCTGAGCTGTATACCGGTTTACGGCTTACCGCTG 988
381 heser...ThrSerCysLeuThrValLeuGly... 390
989 AGTGTGAATTTGTCCGCACTGATCGCCAAAGTCCAGAGGAGGATGAA 1038
391 ...TrpThrSerAlaSerProAsnTrpAspTrp... 400
1039 GGGAAAGTGCAGGTGTCCGCTCTCAAGGCC...AGGTGA 1076
401 ...AlaLeuSerArgAlaSerLeuLeuLeuSerCysT 412
1077 CATCTCGGCC 1087
412 hTrAlaSerSer 415

```

```

seq_name: /cgn2/6/plodata/2/1aa/6B_COMB.pep:us-09-413-814-78
seq_documentation_block:
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bioecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: patentin Ver. 2.1
; SEQ ID NO 78
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-413-814-78

alignment_scores:
Quality: 134.00 Length: 433
Ratio: 0.784 Gaps: 24
Percent Similarity: 39.492 Percent Identity: 25.173

alignment_block:
US-09-775-693-1 x US-09-413-814-78 ..
Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

107 CCACATTGGCCAGAGAGACTTCGAGAGCCAGAGAGGACTG 156
111 111111111111111111111111111111111111
358 ProSerArgArgArgSerProAlaArgThrProGlyProArg..... 372
157 AAGCTTGGGGCCA...AAAGTGTTCATTGAGAGATGTCAGAGGAGCTT 203
373 .....SerProArgArgArgProSerProSerAlaArgSerProAlaI 387
204 TGTGAGAGAGTTCATCTGGCGGCAATCCAGT.....CCAGGCACGTG 247
387 LertpGlnArgProProArgArgProArgAspArgArgProArgArgLeu 403
248 ATGAGACCGCTTACCTCCGCGACCTCTTGCAGGC.....CTGC 291
404 LeuArgProAlaArgArgLeuAlaArgArgAspProGlyAspValProAs 420
292 ATCGCCCGCAAAAGTGAATGAGCCAGC..... 322
420 pProArgAlaAlaArgArgArgAlaProAlaAlaGlyAlaLeuProGlyA 437
323 ...GGAGGGGGCCCAAGTATGTGTCCACGCGCCACAGAGAGGGA 367
437 LaspolyArgGly.....AlaArgGlyProGlyArgArgArg 449
368 AGCATAGAGTCCGCTTGAAGCTCAGCTGTCTACCTGAGCCGCCAGATA 417
450 AlaProArgArgGly.....AlaAlaAlaArg.ValProProAsp 463
418 AAGTCAATTGCTCCCTGAGAGATGCTGAATTCATCAACCGCTTCAAGG 467

```

```

463 rgalaaspProArgrArgrProAla.....ProAlaValAl 475
468 CCCGATGACCTGATGACTAGCCAAAGCACACGGATTCCTACCCCG 517
476 ArgAlaAlaProAlaValAlArgrArgrSerAlaValAlArgrProGl 492
518 TCACGCCCAAGAACCCGTGACATGATGACACCTCATGCACATCAGC 567
492 yValGln.....HisProv 497
568 TCGAGGCTGGATCTTGAGAACCCCAAGACGCGCTCCAGGCT 617
497 AlArgrAlaAlaArgrArgr.....ProAlaArgrArgr.... 508
618 CTACACGAAGACCCAGACCCCAAGCCCAACACCCCTGACATTC 667
509 .....ArgAlaProProGlnPro.....Va 515
668 TCGAGATGAGTTCAAAAAGGGTCCCTGTGAGGTGACCAACGTCAG 717
515 lArgrArgrAlaAlaProArgrGlyAlaAlaAspArgrValGlnArgrGln 532
718 GATGGCACCCACCCACAGAC.....CTCCTTGAGCTCTTCATGTA 761
532 lArgrArgrAlaValProGlyArgrAlaArgrGlyProGlyAlaValPro 548
762 GAA...CGAAGTCCGGGGCAAGCATGGCTGGGCG..... 794
549 AspValArgrProArgrAlaProArgrGlyArgrGlyProArgrArgrGlyAla 565
795 ....TATTGACATGCTGGAGAACCGCTTCATGGAATGAAATCCGAG 840
565 uGlyAlaArgrProArgrGlyAlaProArgrAlaValAlaProGlnProArgr 582
841 ATCAGACAGACCCAGACGACCATCTTACCATGCTCATTTAGACAT 890
582 lAlaAlaProProArgrAlaGlyHisProGlnAlaProArgrArgrAlaArgrHis 598
891 CGAGGC.....CT 898
599 ArgAlaArgrAlaProArgrArgrAlaArgrAlaValAlaArgrArgr 615
899 TCACCATGAGACCGGAGAGTCCCAAAATCAACAGCGCTGGCTGAAA 948
615 uArgrArgrAlaAlaArgrAlaAlaArgrArgrArgrAlaProAla 632
949 TTGCTGAGCTGTGTATACCGGTTAGGCGCTAGCCTGAGTGAATT 998
632 heProAlaAlaGly..... 636
999 TGTCCGCCACATGATGCCAAGTCCACAGACG.....AGTGAAG 1039
637 .....AlaProGlyAlaValAlaArgrArgrAlaArgr 646
1040 GGAAGATGACAGTGTCTCT...CCTCAAGGCGCAGGTGTACAT..... 1079
646 gSerProAlaGlyAlaAlaValAlaArgrArgrGlyProGlyAlaAlaPro 663
1080 .....CCTCGCGCGGGA 1091
663 AlLeuAspAspGluAlaLeuGlyArgrAlaAlaProGlyProAlaGlyAla 679
1092 GTCCCCACTGTC.....TCTCTCAATGAGAGCTGGTAGAGA 1129
680 ArgrProArgrAlaGlyAlaAlaValAlaAspValAlaArgrGlyAlaAlaGlu 696
1130 TGAACGTGACAGGCTGATTATGACCAACTGATGCCACGGTTCAATCAAC 1179
696 oArgrArgrArgrGly.....ArgAlaHis.... 703
1180 ATCAATTCCTCAGGCTGAAGGATATCTCTCTCCAGACGACGAT 1226

```

```

704 .....ProProAspGlnGly 708
seq_name: /cgn2_6/ptodata/2/laa/6B_COMB.pep:US-08-483-533-41
seq_documentation_block:
; Sequence 41, Application US/08483533
; Patent No. 6172047
; GENERAL INFORMATION:
; APPLICANT: Rolzman, Bernard
; TITLE OF INVENTION: Method for Treating Tumorigenic
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-MAR-95
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-95
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/861,233
; FILING DATE: 31-MAR-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-483-533-41

alignment_scores:
Quality: 131.50 Length: 343
Percent Similarity: 1.035 Gaps: 19
Percent Identity: 37.026 Percent Identity: 24.490

alignment_block:
US-09-775-693-1 x US-08-483-533-41 ..
Align seg 1/1 to: US-08-483-533-41 from: 1 to: 355

206 TGAAGAGTTTCATCTGGCCGCCATCTCCAGCGCATGTATAGAGAC 255
111 .....
82 TrpProAspSerProProGlnSerAlaProGln.....AlaAr 95
256 CGCTACTCTCTGGCACCTCTCTTGGCAGGCGCTGATCGCCGCAACA 305
111 .....
95 gProTrpAlaAlaAlaProArgrProGlnGlyProHisArgrProAla.... 110
306 AGTGAATTCGCCAGCGGAGGCGCCAAAGATATGTCTCCAGCGCCCA 355
111 .....

```



```

111 ..TrrpalaaraglyAlaGlyLeuthrPro.....ProThrProPro 123
356 CAGGAAAGGGAGACGATCAGGTCCGGTTTGAGCTCAGCTGCTACTCACTG 405
124 ArgAla..... 125
406 GCGCCCGCAGATAAAGTCATTGCTCCCTGGAGATGCCCTGAATTCTACAA 455
126 .ProserAlaPheArGAlaSerProserAlaCysAlaSer..... 139
456 CCGGTTCAGAGGCCCGCATGACCTGATGAGTACGCAAGAACACAGCGGA 505
139 ..... 139
506 TTCCCATCCGGTCTACCTCCAAAGCCGTGAGCATGATGAGAACCTC 555
140 .....ProArGserThrTrpAlaGlyAlaCysAla 148
556 ATGCACATCAGCTACGAGGTGGAATCTTGAGAACCCCAAGAACCAAGC 605
148 acysaspAla.....ArgAlaGlyArGlyArG 158
606 GCCTCCAGGTCTCTACAGAGACCCAGACCCAGAACCCCGCCACCA 655
158 rg.....SerProProArGProProArGProProArGProProArG 171
656 CCCCTGACATTCTCGATCGATTCGATCAAAAAAGGGTCCCTGTGAAGGTG 705
172 ProProArGProProArGlyCysAlaSerArG..... 182
706 ACCAAGTCAGAGATGACCA.....CCCAAGACCTCTT 743
183 .ProThrSerGlyCysAlaThrTrpSerGlyProArGProAlaTr 199
744 GGAGCTCTTCATGTA..... 758
199 rPArGAlaAlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySe 215
759 .....CCTGAACGAAGTCGCGGACAGATGCGCGTG 789
215 rGlyAlaGlyTrpArGArGProArGArGSerSerGlyArGAlaTrp...G 231
790 GCGCGTATTGACATCGTGAGAACCGCTTCATTTGAATGAAGTCCCGAGG 839
231 lYPro.....ArGProValProGlyProTrpProAlaGluProAla 244
840 TATTCAGAGACCCAGCGACGACCATCTTACCATGCTCATTTAGACA 889
245 ..ArGArGThrArGSerAsnValThrPro...GluAlaAlaTrpValPh 259
890 TCGAGGCTTCACCATGACCGGGAAGTCGCAAAATCAACAGAGCGTG 939
259 eArGly.....AlaProGlySerSerAlaProSerArGSerProG 273
940 GCGTGAATTTGCTGAGCTGCTGATACCGGTTTACGGCCCTAGCCCTGA 989
273 lArGArGTrpGlnGluProArGlyLeuThrLeuGlyAlaSerProPro 289
990 GGTGAAATTTGCGCCACTGCATGCCCAAGTCCAGAGCGAGTGAAG 1039
290 SerGlnGlyGlyProProArGlyArGAspTrpPro...lLeuGlylAr 305
1040 GGAAGTGCAGGTGTCGT..... 1058
305 gGlnArGlyGlyAlaArGProThrSerValArGArGValPheGlyAla 322
1059 .....CCTCAAGGGCCAGGTGATACATCTCTCG 1085
322 rProlleGlyArGgluLeuProProAsnGlyProGly...ArGProLeu 337
1086 CCGGAGATCCCGCAGCTGCTCTACAA 1112
338 ProGlyMetValIleLeuLysAsnLeuGln 346

```

```

seq_name: /cgn2/6/ptodata/2/1aa/PCTUS.COMB.pep:PCT-US91-06532-3
seq_documentation_block:
: Sequence 3, Application PC/TUS9106532
: GENERAL INFORMATION:
: APPLICANT: Roizman, Bernard
: TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Marshall, O'Toole, Gerstein, Murray &
: STREET: Two First National Plaza Suite 2100
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentia Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US91/06532
: FILING DATE: 19910910
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Gruber, Lewis S.
: REGISTRATION NUMBER: 30,060
: REFERENCE/DOCKET NUMBER: 27373/8235
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/346-5750
: TELEFAX: 312/984-9740
: TELEX: 25-3856
: INFORMATION FOR SEQ. ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 355 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US91-06532-3

alignment_scores:
Quality: 131.50 Length: 343
Ratio: 1.035 Gaps: 19
Percent Similarity: 37.026 Percent Identity: 24.490

alignment_block:
US-09-775-693-1 x PCT-US91-06532-3 ..
Align seg 1/1 to: PCT-US91-06532-3 from: 1 to: 355

206 TCGAGAGATTCATCTGGCGCGCATCCAGTCGACGACATGTATGAGAC 255
111 .....AlaAr 95
82 TrpProAspSerProProProGluSerAlaProGlu..... 95
256 GCGTACCTCTGCGACCTCTCTTGCAGGCGCTGCATCGCCGCAACA 305
111 ..... 110
95 gProThrAlaAlaAlaProArGProProGlyProHIsArGProAla.... 110
306 AGTGAATAATCGCCAGCGGAGGGGCGCAAGTATGTGTCCACAGCGCCA 355
111 ..... 111
111 ..TrrpalaaraglyAlaGlyLeuthrPro.....ProThrProPro 123
356 CAGGAAAGGGAGACATCAGGTCCGGTTTGAGCTCAGCTGCTACTCACTG 405
124 ArgAla..... 125
406 GCGCCCGCAGATAAAGTCATTGCTCCCTGGAGATGCCCTGAATTCTACAA 455
126 .ProserAlaPheArGAlaSerProserAlaCysAlaSer..... 139

```

```

456 CCGGTTCAAGGCGCGCAATGACCTGATGAGTACCAAGCAACACGCGGA 505
139 ..... 139
506 TTCCATCCCGGCTACTCCCAAGAACCCGTGAGATGATGAACTTC 555
140 .....ProArgSerThrTrpArgLysal 148
556 ATGACATCAGCTACGAGCGGTGGAACTCTGGAGAACCCCAAGCAAGC 605
148 acysaspala.....ArgAlaGlyArgGlyArg 158
606 GCGTCCAGTCTCTACAGAAAGCCAGCAAGCCCAAGCCCAACCA 655
158 rg.....SerProArgProArgProArgProArgProArg 171
656 CCCCTGACATTCGAGATGAGTCAAAAAGGGTCCCTGTGAAGTG 705
172 ProArgProArgProArgLysalaserArg..... 182
706 ACCAGCTCAAGATGCGACCA.....CCACCAAGCTCTCT 743
183 ..ProTrpSerGlyCysAlaThrTrpTrpSerGlyProArgProAla 199
744 GGAGCTCTTCATGTA..... 758
199 rprargalala..AlaArgGlyProAlaSerGlyProThrGlyLeuGlyse 215
759 .....CTGACAGAAAGTCCGGGCAAGCATGAGCTG 789
215 rglalaglyTrpArgProArgProArgSerSerGlyArgAlaTrp...G 231
790 GCGCGATTGACATCGTGGAGAACCGCTTCATGGAATGAAGTCCCGAG 839
231 lpro.....ArgProValProGlyProTrpProAlaGluProAla 244
840 TATCTACGAGACCCAGCAGACGACATCCTTACATGCTCATTTAGACA 889
245 ..ArgArgThrArgSerAsnValThrPro...GluAlaAlaTrpValh 259
890 TCAGAGCCTTCACCATGAGACCGGGAAGTCCGCAAAATCAACAGAGCTG 939
259 eaGlyGly.....AlaProGlySerSerAlaProSerArgSerProG 273
940 GCGTTGAATTGCTGAGCTGTGATACCGGTTACGGCTCAGCCCTCA 989
273 lnaArgArgTrpGlnGluProArgGlyLeuGlyThrLeuGlyAlaSerPro 289
990 GTGTGAATTGTCCGCACTGATCCCAAGTCCCAAGAGCGAGTGAAG 1039
290 SerGlnGlyGlyProProArgGlyArgAspTrpPro...IleGlyGlyArg 305
1040 GGAAGTGCAGGTGTCCT..... 1058
305 gelnArgGlyGlyAlaArgProThrSerValArgArgValPheGlyAla 322
1059 .....CCTCAAGGCGCAGGTGATACCTCGG 1085
322 rglProIleGlyArgGluLeuProProAsnGlyProGly...ArgProLeu 337
1086 CCGGAGTCCCACTGCTCTCTACAA 1112
338 ProGlyMetValIleLysAsnLeuGln 346
seq_name: /cgn2_6/ptodata/2/lae/5a_COMB.pep:US-08-642-255-72
seq_documentation_block:
; Sequence 72, Application: US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.

```

```

TITLE OF INVENTION: High Molecular Weight Collagen-Like
NUMBER OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A5556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 1065 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-72

alignment_scores:
Quality: 130.00 Length: 397
Ratio: 0.818 Gaps: 27
Percent Similarity: 40.050 Percent Identity: 28.463

alignment_block:
US-09-775-693-1 x US-08-642-255-72 ..
Align seg 1/1 to: US-08-642-255-72 from: 1 to: 1065
105 GGCCAGACATTGGCCAGAGAAAGACTTGAAGAGCCAGGAAGAGCAC 154
277 GLYLAHISGLYPROLALAGLYPRO...LYSGLYLAHISGLYPROALA.. 291
155 TGAAGCTTGGGCCAAAGAGGTTCATTTGAGATGT.....C 192
292 .....GLYPROLYSGLYLAHISGLYPROLALAGLYPROLYSGLYA 305
193 AGCAGGAGTTTGGAGAGATTATGTCGCCGCCATCCAGTCCAGCC 242
305 lagnGlyProAlaGlyProGlyGlySerArgGlyAspProGlyPro 321
243 ACTGTATGAGACCGCTACCTCTGGCA.....CCTCTCT 279
322 GLYLAHISGLYPROLALAGLYPROGlyGlySerArgGlyAspProGly 338
280 GCCAGGCCCTGATCGCCGCAAAACAAGTGAATGCCCGCCAGGAGG 329
338 oProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGlyAsp 355
330 GGCA.....AGTATGTCTCCAGGCGCA...CAGAAAGGGAAG 370
355 lProProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGlyAsp 371
371 ATCAGTCCGTTGAGCTCAGCTGCTACCTGCGCCCGCCAGATMAAG 420

```

```

372 .ProGlyProProGlyAlaGln ..... GlyProAlaGlyProG 384
421 GTCAATTGCTCCCTGGAGAGATGCTGAATTCTTACAACGG... GTTCAAGG 467
11 :::: 111 ::::|||||
384 TlyGlySerArgGlyAspProGly ..... ProProGlyAlaGlnGly 397
468 CCCCAATGACCTGTATGGAGTACGGCAAGAACACACAGCGGATTTCCCATCCCGG 517
111 111 111 111 111 111
398 ProAlaGlyProGlyGlySerArg ..... GlyAspProGlyProTr 411
518 TCACATCCCAAGAACCCGCTGAGCATGTGATGAAACCTCATGCAATCCACG 567
:::111 111::: 111
411 OGlyAlaGlnGlyProAlaGlyProGly ..... GlyS 422
568 TACAGAGCTGGAAATCCGAGAACACCC ..... CAAGAACCAAGCGGCTCCG 611
||||| 111 111 111 111 111
422 eArgGly... AspProGlyProProGlyAlaGlnGlyProAlaGlyPro 437
612 AGGTCTGTACACGAAAGACCCAGGACCC ..... AGCCAAAGCCCCCAACA 655
::: 111 111 111 111 111 111
438 GlyTlySerArgGlyAspProGlyProProGlyAlaGlnGlyProAlaGly 454
656 CCCCTGACATTTCTGA ..... GATCGAATTCAAAAAGGGGTCCCT 696
111 111 111 111 111 111
454 yProGlyTlySerArgGlyAspProGlyProProGlyAlaGlnGlyProA 471
697 GTGAAGGTGACCAACGTCAGAGATGGACACCCACACGCTTCCTTGA 746
GlyArgProGlyGlySerArgGlyAspProGlyProProGlyAlaGlnGly 487
747 GCTCTTCATGTACTGAAACGAAGTCGCGG ..... 776
111 111 111 111 111
488 ProAlaGly... ProGlyTlySerArgGlyAspProGlyProProGlyAl 503
777 .CAAGCATGGGCGGCGGCTATGTGACATGTGGAGAACCGCTTCATTGGA 825
||||| 111 111 111 111 111
503 aGlnGlyProAlaGlyProGlyGlySerArgGlyAspPro ..... 516
826 ATGAAGTCCCGAGGATATGATACGAGAACCCAGCGACGACATCCTTTACCA 875
111 111 111 111 111
517 ..... GlyProProGlyAlaGlnGlyProAlaGly 526
876 TGCATATTTAAGCAT ..... CGAGGCGCTTAC ..... 902
::: 111
527 ProTlyGlyAlaGlnGlyProAlaGlyProTlyGlyAlaGlnGlyProAl 543
903 ..... CATGACCCGGGAAGTGGCAAAATCAACAAGCC 936
111 111 111 111 111 111
543 aGlyProTlyGlyAlaGlnGlyProAlaGlyProTlyGlyAlaGlnGly 560
937 CTGGGCTTGAATTTGCTGAGCTGGGTATACCGGTTTACGGGCTAACCC 986
111 111 111 111 111
560 To ..... AlaGly ..... 562
987 TGAAGTGTGAATTTGTCGCGACATGTCGCAAGTCCAGAGAGCGAGTGG 1036
111 111 111 111 111
563 ..... ProGlyGlySerArgGlyAspProGlyProProGly 574
1037 AAGGCAAGTGCAGAGTGCCTCGCTCAAGGCGCA... GGTTTCAATCTTC 10833
1 :::: 111 111 111 111 111
574 yAlaGlnGly ..... ProAlaGlyProGlyTlySerArgGlyA 587
1084 GGCCGGGAGTCCCGACCTGTCTCTCAATAGAGAGAGCTGGAGCATGAA 11333
||||| 111 ::: 111 111 111
587 sPrProGlyProPro ..... GlyAlaGlnGlyProAlaGlyProGly.Gl 601
1134 CGTGACAGGTGATTAAGCAACATGATGCCACCGGG 1170
: :::11111 111 111
601 ySerArgGlyAspProGlyProProGlyAlaGlnGly 613

```

Sequence 2, Application US/07945283
Patent No. 5352596

GENERAL INFORMATION:

APPLICANT: Cheung, Andrew K.
APPLICANT: Wesley, Ronald D.
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
TITLE OF INVENTION: Involving The Ep0 and Lf1 Genes
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis P. Ribando
STREET: 1815 No. 5352596th University Street
CITY: Peoria
STATE: IL
COUNTRY: USA
ZIP: 61604

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATE:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4011 ext. 513
TELEFAX: 309-685-4128

INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1958 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-945-283-2

```

alignment_scores:
  Quality: 128.00      Length: 413
  Ratio: 0.877        Gaps: 20
  Percent Similarity: 35.351  Percent Identity: 21.792

alignment_block:
  US-09-775-693-1 x US-07-945-283-2 ..

Align seg 1/1  to: US-07-945-283-2  from: 1  to: 1958

221 GCGCGGCATCCAGTCCAGCCGACCTGATGAGACGCGCTACGTCGCGG 270
    ||| ||||| ||| :|||
297 GlycerProPterSerHis..... 304

271 ACCTCTTCGAGCGCCCTGCATCGCCGCAACAGTGAATCGGCCA 320
    ||| ||||| :||| :||| :||| :|||
305 .ProGlyProProGlyProSerGlyProSerThrSerHisInAspPro 321

321 GCGGAGGCGGCCAAGATGTGTCCACAGCGGCCACAGAAAGGGAAC 370
    ||||| ||||| :||| :||| :||| :|||
321 ro.....GlyProProThrProSerThrSerHisHisHisHisHis 335

371 ATCAGGTCCGGTTTGAGCTCAGCACTGCTACTCAGCGCCCGCCAGATAAG 420
    :||| :||| ||| :|||
335 sGlnGlyProPro.....ThrSerPro..... 342

421 GTCATTGCTCCCTGAGAGATGCTGTAATTTCATACCG..GTTCAAAGG 467
    :||| :||| :||| :||| :||| :|||
443 .....ArgProSerThrSerSer 348

468 CCGCAATGACCTATGAGCTATACGAAAGCAACACAGCGATTCCTCCCG 517
    ||| ||| ||| :||| :||| :|||
349 HisGlnAspProProGlyGlyGlyProProSerAlaGluThrHisHisHis 365

```

```

seq_name: /cgn2_6/plodata/2/iaa/5B-COMB.pep:US-08-735-041A-2
592  argAlaGlaGlyAlaIleProArgSerGlnIaGluGln 603
      :::::::::::::::::::: ||| ::::|
seq_documentation_block:
; Sequence 2, Application US/08735041A
; Patent No. 5914251
; GENERAL INFORMATION:
; APPLICANT: Parrell, Catherine L.
; APPLICANT: Martin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1640 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735,041A
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 539 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-735-041A-2

alignment_scores:
      Quality: 127.00      Length: 410
      Ratio: 0.760      Gaps: 26
      Percent Similarity: 40.732      Percent Identity: 25.366

alignment_block:
US-09-775-693-1 x US-08-735-041A-2 ..

Align seg 1/1 to: US-08-735-041A-2 from: 1 to: 539

  98 CCGATGCGCCACAACTTGGCC.....AGAAAGAAAGCTTGGAG 135
      ||| ||| ||||| ||| |||
  185 ProSerHisProAlaGlyLeuAlaIleLeuLeuIleuArgLeuProArgLeuAr 201
      ::::| ::::| ::::| ::::|
  136 GAAGCAGGAGGAAGGACGACGACGACGCTTGGGGCCAAAAGGCGTTCATTGA 165
      ||||| ||| ::::| ||| ::::|
  201 GlySerProLeuArgAlaAspAlaIleAlaIleSerMetHisArgValGly...S 217
      ::::| ::::| ::::| ||||| |||||
  186 GGAATGTACGACGAGGAGTTGTGGAGGAGTTCATCTGGCCGG..... 226
      ::::| ::::| ::::| ||||| |||||
  217 erProAsnAlaIleValLeuIleuTrpSerSerHisAsnArgValAsn 233
      ::::| ::::| ::::| ::::|
  227 .....CGATTCCAGTCCACGACACTG 246
      ::::| ::::| ::::| ::::|
  234 AlaArgLeuGlnValProProAlaIleArgThrProSerSerPro...ArgCy 249
      ::::| ::::| ::::| ::::|

```

```

247 TATG.....AGACCGCTACTCTCTGG 269
      ||| :|||
249 sSerGlyHisProValAsnProValLeuProAlaThrMetLysAlaIlePro 265
      ||| :|||
270 CACCTCTCTTGGCAGGCGCTGATCCGCGGCAACAACTGGAATTCGCC 319
      ||| :|||
276 .....MetCysProCysGlyThrTrpLysProPro 275
      ||| :|||
320 AGCGGGAGGGGGCCAAATATGTGTCCCAAGCGCGCCACAGGAAAGGGGAC 369
      ||| :|||
276 Ser.....ThrsSerArgProThrSerProGln..... 285
      ||| :|||
370 GATCAGGTCCGGTTGAGCTCAGCTGCTACTGCTGCGCCCGCCAGATAAA 419
      ||| :|||
286 .....AlaThrSerSerTrpThrSerLeuGlnL 295
      ||| :|||
420 GGTCAATTGCTCCCTGAGATGCGCTGAATTCTACAAACGGTTCAGAGGCC 469
      ||| :|||
295 euGlnLeuProGlnGlyCysAlaGlnCysAlaAlaIleProGluLeu 311
      ||| :|||
470 GCATGACCTGATGAGTAGCGCAACAGCAGGAGTCCATCCCGGTC 519
      ||| :|||
312 AlaMetGlyAlaLeuGlnLeuGlnSerArgAsnSerThrLeuAspProGln 328
      ||| :|||
520 ACTCCCAAGAACCCGTGGAGCATGATGAGAACCTCATGCATCAGCTA 569
      ||| :|||
328 YLysProGln.....MetLysSerProThrAsnTrpT 340
      ||| :|||
570 CGAGGCTGGAAATCTCTGAGAACCCCAAGCAAGCGCTCCAGCTCTCT 619
      ||| :|||
340 hr.ProHisValProAlaGlnGlyProGlnAlaSerArgProProLysLe 356
      ||| :|||
620 ACAGAGAGCCCAAGA.....CCAGACCAAGCCCAAGCAAGC 657
      ||| :|||
356 whis.....ProGlyLeuArgAlaAlaProGlnGlnLPro.....P 369
      ||| :|||
658 CCTGACATTCGAGATCGAATTCAGAAAAAGGGTCCCTGAGAGGTG 707
      ||| :|||
369 roGlnHisMetAlaAsp...ValGlnArgAsn.....GlnGlnAsp 381
      ||| :|||
708 CAAGCTCAAGGA...TGGCACCAACCCAGACCTCTTGGAGCTTCA 754
      ||| :|||
382 GlnProLeuGlnGlnTrpHisLeuArgSerGlnTrpGlnGly...LeuH 397
      ||| :|||
755 T.....GTACCTGAACGAATCGCGGGGCAACGATGCG 786
      ||| :|||
397 scYstrLeuSerProGlyLeuArgArgThrAlaSerGlyAlaLeuTrpA 414
      ||| :|||
787 GTGG.....GCCGATTCGATCGTGGAGAACCGCTTCATTTG 823
      ||| :|||
414 rSerGlyAlaTrpAlaAlaAlaProSerSerTrpSerThr..... 427
      ||| :|||
824 GAATGAAGTCCGAGGATATCTACGAGACCCAGCAGGACCATCTTAC 873
      ||| :|||
428 .....SerLeuArgProAlaGlnGlyGlyProSer..... 436
      ||| :|||
874 CATGCTCATTTAGACATCGAGGCTTCACCATGACCGGGAAGTGCAG 923
      ||| :|||
437 .....TrpThrGlnArg..... 440
      ||| :|||
924 AATCAACAAGGCGCTGGCTTGAATTTG..... 952
      ||| :|||
441 .....GlnGlnTrpLeuGlnValLeuGlnGlyGlyLysSerTrpLeu 454
      ||| :|||
953 .....CTGAGCTGGTGTATACCGGTTTAC..... 976
      ||| :|||
455 AspIleSerLeuCysValGlnGlyLeuTrpProCysProSerTrpAlaCysTr 471
      ||| :|||
977 .....GGCTACGCGCTGAGCTGTGAAT...TTCCTCG 1004
      ||| :|||
471 pHISValHisLeuLeuProGlnGlnAspLysAlaLeuAsnArgMetLeuA 488

```

```

1005 CCATCTGATCGCCAGATCCAGAGCGAGTGGAGGAAAGTCAAGTGT 1054
      ||| :|||
488 lArThrLeuGlnProGlnLProProGlnGlnAlaGlnGlnGlnAlaAla 504
      ||| :|||
1055 CCGTCTCAAGGGCCAGGTGATCATCC.....TCGGCGGAGATCC 1095
      ||| :|||
505 lIleSerArgHisLeuLysProProAspProIleProSerProProThrPr 521
      ||| :|||
1096 CCACCTCTCTCTACAAATGAGGAGCTGG 1123
      ||| :|||
521 ocYssSerLeuSerGlyLeuGlnValTrp 530

```

seq_name: /cgn2_6/plodata/2/iae/6a_COMB.pep:us-09-190-476B-2

```

seq_documentation_block:
/ Sequence 2, Application US/09190476B
/ Patent No. 6025204
/ GENERAL INFORMATION:
/ APPLICANT: Fartell, Catherine L.
/ APPLICANT: Martin, Francis H.
/ APPLICANT: Yabkowitz, Rachel
/ TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH
/ TITLE OF INVENTION: FACTOR
/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Amgen Inc.
/ STREET: 1840 De Havilland Drive
/ CITY: Thousand Oaks
/ STATE: California
/ COUNTRY: USA
/ ZIP: 91320-1789
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/190,476B
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/735,041
/ FILING DATE: 22-OCT-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Mazza, Richard J.
/ REGISTRATION NUMBER: 27,657
/ REFERENCE/DOCKET NUMBER: A-414
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 805,447,4112
/ TELEFAX: 805,499,6751
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 539 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-190-476B-2

alignment_scores:
Quality: 127.00 Length: 410
Ratio: 0.760 Gaps: 26
Percent Similarity: 40.732 Percent Identity: 25.366

alignment_block:
US-09-775-693-1 x US-09-190-476B-2 ..
Align seg 1/1 to: US-09-190-476B-2 from: 1 to: 539
      98 CCTATCTGCCCAACATTGGCC.....AGAGGAGAGACTTTCGAG 135
      ||| ||| ||| |||
      185 ProSerHisProAlaGlnAlaAlaLeuLeuArgLeuProArgLeuArg 201

```

```

136 GAAGCCAGAGAGGACCTGAGCTTGGGCGCCAAAGGCTTCAATTA 185
    ||||| |||
201 GlyProLeuArgAlaAspAlaAlaSerMetHisArgValGly...s 217
186 GGATGTCAGCAGGAGCTTTGTGAGAGATTTCATCTGGCCGG... 226
    ::::: |||||
217 ePrAsnAlaAlaValLeuTrpLeuTrpSerHisAsnArgValAsn 233
227 .....
234 AlaArgLeuGlnValProProAlaArgThrProSerSerPro...ArgCys 249
247 TATG.....AGACCGCTACCTCTGGG 269
249 sSerGlyHisProValAsnPhenValLeuProAlaThrMetAsnAlaTrp. 265
270 CACCTCTCTGGCCAGGCGCTGCATCGCCCGCAACAGTGGAAATCGCC 319
266 .....MetCysProCysGlyThrTrpIleProPro 275
320 AGCGGAGGCGGCCAAGTATGTCTCCAGCGCGCCACAGAGAAAGGAGAC 369
    ||| |||
276 Ser.....ThrsSerArgProThrSerProGln..... 285
370 GATCAGGTCGGGTTGAGCTCAGCTGACTGCTACCTGCGCCCGCAGATAAA 419
286 .....AlaThrSerSerTrpThrSerLeuGln 295
420 GGTCAATGCTCCCTGAGATGCCGGAATTTACACCGGTTACAGGCGC 469
    ::| ||||| |||||
295 euGlyGlnLeuProGlnGlyCysAlaGlyCysAlaAlaAlaProGlnLeu 311
470 GCATGACCTGATGAGTACGCAAGACACAGGAGATTCCATCCCGCTG 519
    |||||
312 AlaMetGlyAlaLeuGlnLeuGlnSerArgAsnSerThrLeuAsnProGln 328
520 ACTCCAGAGACCGGTGAGATGATGAGAACTCATGCACTACAGCTA 569
    : ||||| |||
328 LysProGlu.....MetMetLysSerProThrAsnThrT 340
570 CGAGGCTGGGAATCTGAGAAACCCAGAACACGCGCTCGAGGCTCTG 619
    ||| ||||| |||
340 hr.ProHisValProAlaGlnGlyProGlnAlaSerArgProProLysLe 356
620 ACACGAGACCCAGGA.....CCGACCCAAAGCCCGCAACAGCC 657
    |||||
356 uHis.....ProGlyLeuArgAlaAlaProGlnGlnPro.....P 369
658 CCGACACTTCTGAGATCGATTCAAAAAGGGTCCCTGTGAGAGGTGAC 707
    ||| |||
369 roGlnHisMetAlaAsp...ValGlnArgAsn.....GlnGlnAsp 381
708 CAACGTCGAAGA...TGSCACCAACCCACAGACTCTTGGAGCTCTTCA 754
    ::| |||||
382 GluProLeuGlnGlnTrpHisLeuArgSerGlnThrGlnGly...LeuHis 397
755 T.....GTACTGAACGAGAGTCGCGGAGCAGATGCG 786
397 sCysTrpLeuSerProGlyLeuArgArgThrAlaSerGlyAlaLeuTrpA 414
787 GTGG.....GCCGATTTGACATCGCTGGAGAACCGCTTCATG 823
    |||
414 rGserGlyAlaTrpAlaAlaAlaProSerSerTrpSerThr..... 427
824 GAATGAAGTCCGAGATCTCTACGAGACCCAGAGCGACGACATCTTAC 873
    ||| |||||
428 .....SerLeuArgProAlaGlyGlyProSer..... 436
874 CATGCTCATTTAGACATGAGGCTTACCATGAGACCGGGAATGCCAA 923
    |||||
437 .....TrpThrGlyArg..... 440

```

```

924 AATCAACAGAGCGCTGGCTTGAATTTG..... 952
441 .....GlyGlnTrpLeuGlnValLeuGlyGlyPheSerTrpLeu 454
953 ...CTGAGCTGTGTATACCGGTTTAC..... 976
455 AspIleSerLeuGlyValGlyLeuTrpProCysProSerTrpAlaCysTr 471
977 .....GCCCTAGCCCTGAGTGTGAAT...TTGTCCG 1004
471 pHISValHisLeuLeuProGlyGlnAspLysAlaLeuAsnArgMetLeuA 488
1005 CCACCTGATCGCCAAAGTCCAGAGAGAGTGGAGAGAAAGTGCAGCTGT 1054
    ||||| ::|||
488 lArThrLeuGlnProGlnProProGlyGlnGlnAlaGlyGlyAla 504
1055 CCGTCTCTAAGGCGCAGGTGTACATCC.....TCGCGCGGAGTCC 1095
    ||| ::|
505 lIleSerArgHisLeuLysProProAspProIleProSerProProThrPr 521
1096 CCACGTGTCTCTACATGAGAGAGCTGC 1123
521 ocysSerLeuSerGlyLeuGlnValTrp 530

```

seq_name: /cgn2_6/prodata/2/1aa/6A_COMB.pep:us-09-190-889A-2

seq_documentation_block:

Sequence 2, Application US/09190889A

Patent No. 6075008

GENERAL INFORMATION:

APPLICANT: Farrell, Catherine L.

APPLICANT: Martin, Francis H.

APPLICANT: Yabkowitz, Rachel

TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSER: Amgen Inc.

STREET: 1840 De Havilland Drive

CITY: Thousand Oaks

STATE: California

COUNTRY: USA

ZIP: 91320-1789

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/190,889A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/735,041

FILING DATE: 22-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mazza, Richard J.

REGISTRATION NUMBER: 27,657

REFERENCE/DOCKET NUMBER: A-414

TELECOMMUNICATION INFORMATION:

TELEPHONE: 805,447,4112

TELEFAX: 805,499,6751

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 539 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-190-889A-2

alignment_scores:

Quality: 127.00 Length: 410
Ratio: 0.760 Gaps: 26
Percent Similarity: 40.732 Percent Identity: 25.366

alignment_block:
US-09-775-693-1 x US-09-190-889A-2 ..

Align seg 1/1 to: US-09-190-889A-2 from: 1 to: 539

```

98 CSTATCTGSCCAATTGGCC.....AGAGGAGAGCTTCGAG 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 ProSerHisProArgLeuAlaLeuLeuLeuArgLeuProAlaGlyLeu 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 GlyProLeuAlaGlyAlaAspAlaAlaSerMetHisArgValGly...S 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 GAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 erProAlaAlaValLeuTrpSerSerHisAsnArgValAsn 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 .....CCATCCAGTCCAGCCAGCTG 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 AlaArgLeuGlnValProProAlaArgTrpProSerSerPro...ArgC 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
247 TATG.....AGAGCCGCTACCTCTGCG 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 sSerGlyHisProValAsnProValLeuProAlaTrpMetAsnAlaTrp 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 CACCTCTCTGSCAGAGCCCTGCATCGCCGCAAGAGAGAGAGAGAG 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....MetCysProCysGlyThrTrpIlyProPro 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
320 AGCGGAGAGGCGCCAGATATGTCTCCAGCGCCAGAGAGAGAGAGAG 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 Ser.....ThrSerArgProThrSerProGln..... 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 GATCAGTCCGCGTTGAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 .....AlaTrpSerSerTrpThrSerLeuGln 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 GGTCTATGTCCTCGAGATGCTGATTTCTACACCGGTTCAGAGGCC 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 euGlyGlnLeuProGlnGlyCysAlaGlyCysAlaAlaProGlnLeu 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
470 GCATGACCTGATGATGATGATGATGATGATGATGATGATGATGAT 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 AlaMetGlyAlaLeuGlnLeuGlnSerArgAsnSerThrLeuAspProG 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 ACTCCCAAGAACCCGTCGATGATGATGATGATGATGATGATGATGAT 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 yLysProGlu.....MetMetLysSerProThrAsnThr 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
570 CGAGGCTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 hrProHisValProAlaGlnGlyProGlnAlaSerArgProProLysLe 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
620 ACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
356 unHis.....ProGlyLeuArgAlaAlaProGlnGlnPro.....P 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 CCGATGATCTCGAGATGATGATGATGATGATGATGATGATGATGAT 707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 roGlnHisMetAlaAsp...ValGlnArgAsn.....GlnGlnAsp 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
708 CAACGTCACAGA...TGACACACACACACACACACCTCTTGAGAGCT 754
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 GlnProLeuGlnGlnTrpHisLeuArgSerGlnThrGlnGly...LeuH 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 T.....GTACCTGACAGAGAGTCCGCGGAGAGAGAGAGAGAG 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 scyStrIleuSerProGlyLeuArgTrpThrAlaSerGlyAlaLeuTrpA 414

```

```

787 GTGG.....GCCGATTTGACATCGTGGAGAGAGAGAGAGAGAGAG 823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 rSerGlyAlaTrpAlaAlaAlaProSerSerTrpSerThr..... 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
824 GAATGAGTCCCGAGGTATCTACAGAGAGAGAGAGAGAGAGAGAGAG 873
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 .....SerLeuArgProAlaGlyGlyProSer..... 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 CATGCTCATTTAGACATCGAGAGCTTACCATGAGAGAGAGAGAGAG 923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 .....TrpThrGlyArg..... 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 AATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 .....GlyGlnTrpLeuGlnValLeuGlnGlyGlyGlyPheSerTr 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
953 .....CTGAGCTGCTGATACCGGTTAC..... 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 AspIleSerLeuGlyValGlyLeuTrpProCysProSerTrpAlaCysTr 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
977 .....GGCCTAGCCCTGAGTGAAT...TTGTCCG 1004
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 PHisValHisLeuLeuProGlyGlnAspLysAlaLeuAsnArgMetLeuA 488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1005 CCACTGCATCGCCAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 LatrLeuGlnProGlnProProGlyGlnGlnAlaGlyGlnGlyAlaAla 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1055 CCGTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1095
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 IleSerArgHisLeuLysProProAspProIleProSerProProThrPr 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1096 CCACTGCTCTCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
521 ocysSerLeuSerGlyLeuGlnValTrp 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

seq.name: /cgn2_6/prodata/2/1aa/6B_COMB.pep:US-09-190-938B-2

seq_documentation_block:
: Sequence 2, Application US/09190938B
: Patent No. 6197939
: GENERAL INFORMATION:
: APPLICANT: Farrell, Catherine L.
: Martin, Francis H.
: Yabkowitz, Rachel
: TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
: FACTOR
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Amgen Inc.
: STREET: 1840 De Havilland Drive
: CITY: Thousand Oaks
: STATE: California
: COUNTRY: USA
: ZIP: 91320-1789
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/190,938B
: FILING DATE: 12-No. 6197939-1998
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Mazza, Richard J.
: REGISTRATION NUMBER: 27,657
: REFERENCE/DOCKET NUMBER: A-414
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 805.447.4112
: TELEFAX: 805.499.6751
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:

```

```

;
; LENGTH: 539 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-190-938B-2

```

alignment_scores:

```

Quality: 127.00 Length: 410
Ratio: 0.760 Gaps: 26
Percent Similarity: 40.732 Percent Identity: 25.366

```

alignment_block:

```

US-09-775-693-1 x US-09-190-938B-2

```

```

Align seg 1/1 to: US-09-190-938B-2 from: 1 to: 539

```

```

98 CCTATGTGGCCCAACATTGGCC.....AGAGGAGACGCTCGAG 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 ProSerHisProArgLeuAlaLeuLeuLeuArgLeuProArgLeu 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GAAGCCAGAGAGAGGCACTGAGCTTGGCCGCAAAAGTGTTCATTGA 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 gTysProLeuArgAlaAspAlaAlaSerMetHisArgValGly...S 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 GATATCGACAGAGGAGTTTGTGAGGAGTTCATCTGGCCGG..... 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 erProAsnAlaValLeuTrpLeuTrpSerSerHisAsnArgValAsn 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 .....CCTACGATCCAGCGCAGCTG 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 AlaArgLeuGlnValProProAlaArgTrpProSerSerPro...Arg 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
247 TATG.....AGAGCCGCTACCTCCGCG 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 sSerGlyHisProValAsnPhenValLeuProAlaTrpMetAsnAlaTrp 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 CACCTCTCTTCCGAGGCCCTGCATGCCGCAAAACAGTGGAAATGCC 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....MetCysProGlySerGlyThrTrpLysProPro 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
320 AGCGGAGGGGGCCCAAGTATGTGCCAGCGCCACAGAGAAAGGGGAG 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 Ser.....ThrSerSerArgProTrpSerProGln..... 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 GATCAGGTCCGGTTGAGCTCACTCTCTACTCAGTGGCCCCCAGATAA 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 .....AlaThrSerSerTrpTrpTrpSerLeuGln 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 GTTCATTGCTCCCTGGAGATGCTGATTTACCAACGCGTTCACAGGCC 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 euGlyGlnLeuProGlnGlyCysAlaGlnCysAlaAlaLeuProGln 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
470 GCATGACCTGATGAGATACGCAAGCAACAGGATTCCTCCCGGCTC 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 AlaMetGlyAlaLeuGlnLeuGlnSerArgAsnSerTrpLeuAspPro 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 ACTCCCAAGAACCCGTCGAGCATGATGATGAACCATGCAATCAGCTA 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 LysProGln.....MetMetLysSerProTrpTrpSerTrp 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
570 CGAGGCTGATCTCGAGAACCCCAAGAACCAAGCGGCTCAGAGTCT 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 hr.ProHisValProAlaGlnGlyProGlnAlaSerArgProPolyS 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
620 ACAGCAAGCCAGCA.....CCACGCAAAAGCCCCACACACC 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
356 whis.....ProGlyLeuArgAlaAlaProGlyGlnGlnPro....P 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 CCTGACATTCGAGATCGAGTTCAAAAGGGGCTCCCTCGAAGGTGAC 707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

369 roGlnHisMetAlaAsp...ValGlnArgAsn.....GluGlnAsp 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
708 CAACGTCACAGA...TGGCACCACCCACGACCTCTTGAGAGCTTCA 754
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 GluProLeuGlnGlyGlnTrpHisLeuArgSerGluTrpGlnGlnGly 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 T.....GTACCTGACAGCAAGTCCGCGGCAAGCAGTGC 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 scYsTrpLeuSerProGlyLeuArgArgTrpAlaSerGlyAlaLeuTrp 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
787 GTGG.....GCCATTGACATCGTCGAGAGAACGCTTCATTG 823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 rSerGlyAlaTrpAlaAlaAlaProSerSerTrpSerThr..... 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
824 GATGAGATCCCGAGGATCTACGAGACCCACGACGACCATCTTTAC 873
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 .....SerLeuArgProAlaGlyGlyProSer..... 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 CATGCTCATTTAGACATCGAGGCCCTTCACATGAGCGGAGAGTCCGA 923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 .....TrpThrGlyArg..... 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 AATCAACACAGCCCGGCTTGAATTTG..... 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 .....GlyGlnTrpLeuGlnValLeuGlnGlyGlyPheSerTrpLeu 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
953 ..CTGAGCTGTGTATACCGGTTTAC..... 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 AspIleSerLeuGlyValGlyLeuTrpProCysProSerTrpAlaCysTr 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
977 .....GCCCTAGCCCTGAGCTGTGAAT...TTGTCCG 1004
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 phiValHisLeuLeuProGlnIleAspLysAlaLeuAsnArgMetLeu 488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1005 CCACTGATCGCCCAAGTCCGAGAGCGAGTGAAGGAAAGTGCAGGT 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 LathrLeuGlnProGlnProGlyGlnAlaGlyGlnGlyAlaAla 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1055 CCGTCTCAAGGGCCAGGTGTACATCC.....TCGGCCGGAGATCC 1095
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 TLeSerTrpHisLeuLysProProAspProIleProSerProProThrPr 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1096 CCACTGTCTCTACATGAGGAGCTGG 1123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
521 ocYsSerLeuSerGlyLeuGlnValTrp 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

seq_name: /cgn2.6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US95-09261-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9509261
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: NAME: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM
; APPLICANT: STREET: 201 West 7th Street
; APPLICANT: CITY: Austin
; APPLICANT: STATE: Texas
; APPLICANT: COUNTRY: United States of America
; APPLICANT: POSTAL CODE: 78701
; APPLICANT: TELEPHONE NO: (512)499-4462
; APPLICANT: TELEFAX: (512)499-4523
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE EXPRESSION OF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09261
; FILING DATE: CONCURRENTLY HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/283,701
; FILING DATE: 01-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: SERTICH, GARY J.
; REGISTRATION NUMBER: 34,430
; REFERENCE/DOCKET NUMBER: UTFC422P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 539 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-09261-2

```

```

alignment_scores:
  Quality: 127.00      Length: 410
  Ratio: 0.760         Gaps: 26
  Percent Similarity: 40.732   Percent Identity: 25.366

```

alignment_block:
US-09-775-693-1 x PCT-US95-09261-2

Align seg 1/1 to: PCT-US95-09261-2 from: 1 to: 539

```

98 CCTATCTGCGCCACATTTGGCC.....AGAAAGCAACTTCGAG 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 PROSERHISPROVALAVALAVALAVALAVALAVALAVALAVALAVALA 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 AGAGCCAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 GLYSERLEUARGALASPRALALALALALALALALALALALALALAL 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 GGATGTCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 ERPROASNAVALAVALAVALAVALAVALAVALAVALAVALAVALA 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 .....CCATCAGTCCAGCGCACTG 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 ALAARGLEUINVALPROVALAVALAVALAVALAVALAVALAVALA 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
247 TATG.....AGACCCGCTACCTCTGGG 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 SSERGLHISPROVALASNAVALAVALAVALAVALAVALAVALAVAL 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 CACCTCTCTGCGAGCGCTGCGAGCGCGCAACAGAGAGAGAGAGAG 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....METSERPROCYSELGLYHTRIPRYSPPROP 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
320 AGCGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 SER.....THRSERARGPROFHISERPROGLN..... 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 GATCAGGTCCGGTTTGAGCTCAGTCTACTCACTGCGCCCGCCAGATA 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 .....ALATRHSERTRIPRHSERLEUINL 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 GGTGATCTCTCCCTGAGAGATGCTGAATTTCTACACCGGTTCAAGGCC 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 EUGLYGHNLEUPROGLNGLYCYSALAGLUCYSALALALALALALALAL 311

```

```

470 GCAATGACCTGATGAGTAGACCAACCAACAGCGGATTCCTCCCGGNC 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 ALATGCLYALALEUINLEUINLEUINLEUINLEUINLEUINLEUIN 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 ACTCCCAAGAACCCGTGACATGATGAGAACCTCATCAGATCAGTAA 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 YLYSPROGLN.....METSELYSERPROTHASHTHT 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
570 CGAGGCTGGAATCTGGAGAACCCCAAGAACCAAGCGCTCCAGTCTCT 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 HR.PROHISVALPROHISVALGLUGLYPROGLUALASERARGPROHIS 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
620 ACACGAGAACCCAGGA.....CCACGCAAGGCCCAACACAC 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
356 UNLS.....PROGLYLEUARGALALALALALALALALALALALAL 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 CCTGACATTTCTGAGATGAGATGAGTCAAAAAGGGTCCCTGAGAGTAC 707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 ROGLHISMETALASP...VALGHNARGASN.....GLUGHNASP 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
708 CAACGTCAAGGA...TGGCACCCACCACACCTCTTGAGAGCTCTTCA 754
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 GLUPROLEUGLYGLNTRPHISLEUARGSERGLNTRHGLNGLY... 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 T.....GTACCTGACAGAGTCCGCGGCAAGCATGCG 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 SCYSTTRPLEUSERPROGLYLEUARGHTRHTRHTRHTRHTRHTRHTR 823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
787 GTGG.....GCCGTAATGACATCGTGAGAACCCGCTCTCTTTG 823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 RYSERGLYALATRPALALALALALALALALALALALALALALALAL 873
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
824 GAATGAATCCCGAGGATTTCTAGAGACCCACGACACATCTCTTAC 873
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 .....SERLEUARGPROALAGLYGLYPROSER..... 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 CATGCTATTATGACATGAGAGGCTTCCACCATGACCGGAGAGTGGCA 923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 .....TRPHNGLYARG..... 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 AATCAACAGAGCCTGGGCTTGAATTTG..... 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 .....GLYINTRPLEUINLEUINLEUINLEUINLEUINLEUIN 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
953 .....CTGAGCTGGTATACCGGTTTAC..... 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 ASPHISERLEUCYSVALGLYLEUITYRPROCYSPROSERTRPALACYSTR 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
977 .....GGCCTAGCCCTGAGTGTGAT...TTGTCG 1004
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 PHISVALHISLEUINLEUINLEUINLEUINLEUINLEUINLEUIN 488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1005 CCACGTGACAGCGCAAGTCCAGAGAGAGAGAGAGAGAGAGAGAG 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 LATHRLEUGINPROGLUPROPROGLYGLUGLUALAGLYGLUALALA 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1055 CCCTCTCTCAAGGCGAGGTGTACATCC.....TGGCGCGGAGTCC 1095
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 ILSERARGHISLEUINLEUINLEUINLEUINLEUINLEUINLEUIN 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1096 CCACGTCTCTCTACATAGAGAGAGTGG 1123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
521 OCYSERLEUSERGLYLEUINLEUINLEUINLEUINLEUINLEUIN 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: /cgn2_6/ptodata/2/1aa/5A-COMB.pep:US-08-642-255-73

```

; seq documentation block:
; Sequence 73, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CARPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like

```

```

? TITLE OF INVENTION: Protein Polymers
? NUMBER OF SEQUENCES: 135
? CORRESPONDENCE ADDRESS:
? ADDRESS: FLEHR, HOHBACH, TEST, ALBERTITTON & HERBERT
? STREET: 4 Embarcadero Center, Suite 3400
? CITY: San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94111-4187
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/642,255
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: ROWLAND, Berttram I.
? REGISTRATION NUMBER: 20,015
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 494-8700
? TELEFAX: (415) 494-8771
? INFORMATION FOR SEQ ID NO: 73:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 633 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-642-255-73

alignment_scores:
  Quality: 126.50      Length: 387
  Ratio: 0.904        Gaps: 25
  Percent Similarity: 36.176  Percent Identity: 26.357

alignment_block:
US-09-775-693-1 x US-08-642-255-73  ..

Align seg 1/1 to: US-08-642-255-73 from: 1 to: 633

105 GCCCAACATTTGCCGAGAACTTCGAGAGCCGAGAGAGAGAGC 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
160 G1yAlaHisG1yProAlaG1yPro..LysG1yAlaHisG1yProAla.. 174
155 TGAAGCTTGGGGCCAAAAGGTGTCATTGAGATGTCAGCAGGAGTTT 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 .....G1yProLysG1yAlaHis.....G1yProAla 183
205 GTGGA.....GGAGTTCATCTGCGCGGCATCCAGTCCAGCGCACTGA 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
183 laG1yProLysG1yAlaHisG1yProAlaG1yPro.....LysG1yAla 197
249 TGAAGACCGCTACCTCTGGGCACTCTCTGCAAGCCCTGCATCGGCC 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 GlnG1yProAlaG1yProGly.....GlySer 224
299 GCAAAACAGTGAATTCG.....C 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 .....GlySerArgG1yAspProG1yProProG1yAlaGlnG1yP 218
319 CAGCGGAGAGGGGCCAAGTATGTCTCCACGCGCCACAGAGAAAGGGAA 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 roAlaG1yProGly.....GlySer 224
369 CGATCAGTCGCGTTTGAAGTCAGCTGCTACTACGCGCCGCCAGATAA 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 ArgG1yAspPro.....G1yProProGly.. 232

```

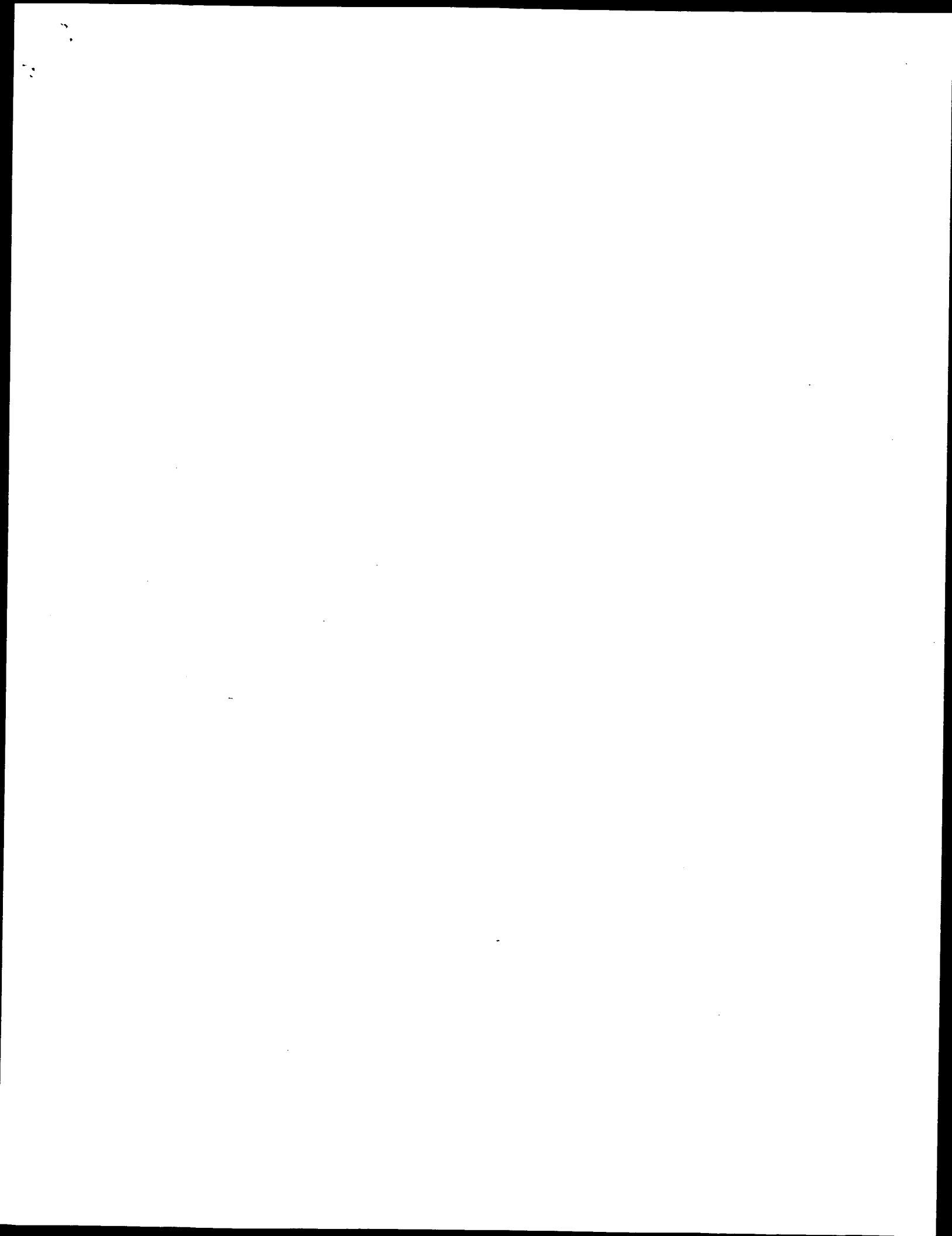
```

419 AGGTCAATGCTCCCTGGAGATGCTGAATCTACACCGGTTCAAGGC 468
233 .....AlaGlnG1yP 236
469 CGCAATGACCTGATGAGTACGCAAGCAGCAGGATTCCTCCATCCGGT 518
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 roAlaG1yProG1yLysArg.....G1yAspProG1yProPro 249
519 CACTCCCAAGAACCCGTGACCATGATGAGAACCTATGACATCACTGCT 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 G1yAlaGlnG1yProAlaG1yProGly.....Glyse 260
569 ACAGGCTGGAATCTGGAGACCC.....CAAGACCAAGGCGCTCA 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 rArgG1y...AspProG1yProProG1yAlaGlnG1yProAlaG1yProG 276
613 GGTCTTACAGAGACCCAGAGACC.....AGCCAAAGCCCAACAC 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 LysG1ySerArgG1yAspProG1yProProG1yAlaGlnG1yProAlaG1y 292
657 CCGTGACATTCGTGAGATCGATTCAAAAAGGGGTCCTGTAAGGTGA 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 ProG1yLysArgG1yAspProG1yProProG1yAlaHisG1yProAl 309
707 CCAACGTCAAGATGACACCCACCCACAGACCTCTGAGCT..... 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 aG1yProLysG1yAlaHisG1yProAlaG1yProLysG1yAlaHisG1yP 326
750 .....CTTCATGTAACCTGAAGCAAGTGGGG...CAA 779
326 roAlaG1yProLysG1yAlaHisG1yProAlaG1yProLysG1yAlaGln 342
780 GCATGCGCTGGCGCTGATTTGACATGTCGAGAGACCGCTTCAATGAATGA 829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 G1yProAlaG1yProG1yLysArgG1yAspPro..... 354
830 AGTCCGAGATTCACGAGACCC.....AGCAGCGACC 864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 ..G1yProProG1yAlaGlnG1yProAlaG1yProG1yLysArgG1yA 371
865 ATTCCTTACATGCTCATTTAGACATGAGCGCTTCACTGAGCGGGA 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 sProPro.....G1yProPro 375
915 AGTGGCAAAATCAACAAGGCTGGGCTTGAATTTGCTGAGCTGTGT 964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
376 G1yAlaGlnG1yProAlaG1yProG1yLysArg..... 386
965 ATACCGGTTACGGCCTAGCCGTGAGTGAATTTGCGCCACTGCATC 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 .....A 387
1015 GCCAAGTCCAGAGAGAGTGAAGGAAAGTGAAGTGTGCTGCTCA 1064
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 rG1yAspProG1yProProG1yAlaGlnG1y.....ProAla 399
1065 GGGCCA...GGTGTACATCTGCGCGGAGAGTCCCACTGTCTCTTACA 1111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 G1yProG1yLysArgG1yAspProG1yProPro.....G1yAlaG1 414
1112 ATGAGAGCTGTGAGCATGAACGTGACGGGTATTTAGACCAACATGAT 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 nG1yProAlaG1yProG1y..GlySerArgG1yAspProG1yProProG1y 430
1162 GCCACCGG 1170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 AlaGlnG1y 433

```

Wed Feb 13 07:35:59 2002

us-09-775-693-1.ra1




```

151 GCACGTGAAGCTGGGGCCAAAAGTCTTCATTCAGAGATGTCAGAGGA 200
51 AlaLeuLysLeuGlyAlaLysValPheIleGluAspValSerArgI 67
201 GTTGTGGAGAGATTTCATTCGGCCGCCATCCAGTCACGACGACGTATG 250
67 uPheValGluGlnPheIleTrpProAlaIleGlnSerSerAlaLeuTyrG 84
251 AGGACCGCTACCTCTGGGCACCTCTTCGACAGGCCCTGCATGCGCGC 300
84 LuAspArgTyrLeuLeuGlnTyrSerLeuAlaArgProCysIleAlaArg 100
301 AACACAGTGAAGTTCGCCAGAGGGGAGGCGCAAGTATGTGCCAGG 350
101 LysGlnValGlnIleAlaGlnArgGlnGlyAlaLysTyrValSerHisG 117
351 CGGCACAGAAAAGGGAGACATCAGTCCGGTTGAGCTCAGCTGCTCT 400
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGlnLeuSerCysTyrS 134
401 CACTGGCCCCCAGATTAAGTCAATTCCTCCCTGGAGATGCTGAATTC 450
134 eLeuAlaArgProGlnIleLysValIleAlaProTyrArgMetProGlnP 150
451 TCAACCGGTTCAAGGGCCGACATGACCTGATGGAGTACGCAAGACAC 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGlnTyrAlaLysGlnH 167
501 CGGGATTCGCATCCCGGTCATCCCAAGAACCCCTGGAGCATGATGAGA 550
167 SGIlyLeuProIleProValThrProLysAsnProTyrSerMetAspGln 184
551 ACCTCATGACATCAGTACGAGGCTGGAATCCGAGGAGACCCCAAGAAC 600
184 snLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLysAsn 200
601 CAACGGCTCCAGTCTCTACAGAGAACCCAGACCCAGCCAAACCCGCC 650
201 GlnAlaProProGlyLeuTyrThrLysThrGlnAspProAlaLysAla 217
651 CAACACCCCTGACATTCGAGATCGAGTTCGAAAAAGGGTCCCTGTGA 700
217 oAsnThrProAspIleLeuGlnIleGlnPheLysGlyValProValL 234
701 AGGTGACCAACGTCAAGATGAGCCACCCAGACCTCCTGGAGCTC 750
234 yValThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGln 250
751 TTCATGTACCTGAACGAGTCCGGGCAAGCATGGCTGGGCGCTATTGA 800
251 PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValGlyArgIleAs 267
801 CATCGTGGGAACCGCTTCATTCGAATGAGTCCGAGGTATCTACGAGA 850
267 rIleValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrGln 284
851 CCCGACGAGGACCATCCCTTACATGCTGCTATTAGACATGAGCGCTTC 900
284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGlnAlaPhe 300
901 ACCATGACCGGGAAGTCCCAAAATCAACAAAGGCTGGGCTTGAATTT 950
301 ThrMetAspArgGlnValArgLysIleLysGlnIleLysLeuLysSph 317
951 TGCTGAGCTGGTATACCGGTTTACGGCTTACGGCTGAGTGTGAATTTG 1000
317 eAlaGlnLeuValTyrThrIleLeuArgProSerProGlnCysGlnPheV 334
1001 TCCGCACTGATCGCAAGTCCAGAGGAGGAGTGAAGGAAGTGCAG 1050
334 aArgHisCysIleAlaLysSerGlnGlnTyrValGlnGlyLysValGln 350
1051 GTTCCTCTCTCAAGGGCCAGGTGTACATTCCTGCGCGGAGTCCCACT 1100

```

```

|||||
351 ValSerValLeuLysGlnValTyrIleLeuGlyArgGlnSerProLe 367
1101 GTCTCTTACAAATGAGAGAGCTGGTGAACATGAGACGTGAGGTATATG 1150
367 uSerLeuTyrAsnGlnGlnLeuValSerMetAsnValGlnGlyAspTyrG 384
1151 AGCCAACTATGCGCACCGGGTTTCATCAACATCAATTCCTCAGCTGAAG 1200
384 LuProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATTCATTCGCTCTCCAGAGCAAGGTCACTGCGCAAA 1236
401 GlnTyrHisArgLeuGlnSerLysValThrAlaLys 412
seq_name: p1r1:AUFRS
seq_documentation_block:
  argininosuccinate synthase (EC 6.3.4.5) - rat
  N:Alternate names: citrulline--aspartate ligase
  C:Species: Rattus norvegicus (Norway rat)
  C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Jun-1999
  C:Accession: S01440
  R:Surv, L.C.; Morris, S.M.; O'Brien, W.E.; Beaudet, A.L.
  Nucleic Acids Res. 16, 9352, 1988
  A:Title: Nucleotide sequence of the cDNA encoding the rat argininosuccinate synthetas
  A:Reference number: S01440; M01D:89016648
  A:Accession: S01440
  A:Molecule type: mRNA
  A:Residues: 1-412 <SUR>
  A:Cross-references: EMBL:X12459; NID:955766; PIDN:CAA30999.1; PID:955767
  C:Comment: This enzyme catalyzes the formation of argininosuccinate from citrulline a
  C:Superfamily: argininosuccinate synthase
  C:Keywords: arginine biosynthesis; homotetramer; ligase; urea cycle
  F:149,153/Binding site: Mg-AMP (Glu, Arg) #status predicted

alignment_scores:
  Quality: 2086.00
  Ratio: 5.113
  Percent Similarity: 99.029
  Percent Identity: 96.845

alignment_block:
  US-09-775-693-1 x AUFRS
  Align seg 1/1 to: AUFRS from: 1 to: 412

1 ATGTCGACGAAAGGCTCCGTTGCTTGGCTACAGTGGCGGCTGACAC 50
|||||
1 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspTh 17
51 CTCGTGCAATCCGCTGCTGAGGAGGAGCAAGGCTATGAGCTCATTCGCT 100
|||||
17 rSerCysIleLeuValIrrPheLysGlnGlnGlyTyrAspValIleAla 34
101 ATCTGGCCACAACTTGGCCAGAAAGACTTCGAGGAAGCCAGAGAG 150
34 yLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGlnAlaArgLysLys 50
151 GCACGTGAAGCTTGGGGCCAAAAGTGTTCATTCAGAGATGTCAGAGGA 200
|||||
51 AlaLeuLysLeuGlyAlaLysValPheIleGlnAspValSerLysG 67
201 GTTGTGGAGAGATTTCATTCGGCCGCCATCCAGTCCAGGACGCTGTATG 250
|||||
67 uPheValGlnGlnPheIleTrpProAlaValGlnSerSerAlaLeuTyrG 84
251 AGGACCGCTACCTCTGGGCACCTCTTCGACAGGCCCTGCATGCGCGC 300
|||||
84 LuAspArgTyrLeuLeuGlnTyrSerLeuAlaArgProCysIleAlaArg 100
301 AACACAGTGAAGTTCGCCAGAGGGGAGGCGCAAGTATGTGCCAGG 350
|||||

```

```

101 LysGlnValGluIleAlaGlnArgGluGlyAlaValSerHisG1 117
351 GGGCCAGGAAGGGGAGACGATGAGTCCGGTTGAGCTGAGCTGCTACT 400
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuThrCysTyrS 134
401 CAGTGGCCCCCAGATAAAGTCATGCTGCTCCCTGGAGAGATCCGTAATC 450
134 eTLeuAlaProGlnIleLeuValIleAlaProTrpArgMetProGluPhe 150
451 TACAAACGGTTCAAGGGCCCAATGAGCTGATGGATCGCAAGCAACA 500
151 TyrAsnArgPheLysGlyArgAsnSpreuMetGluTyrAlaLysGlnH1 167
501 CGGATTCCTCCATCCCGGTCACTCCCAAGACCCGTGAGCATGATGAGA 550
167 sGlyIleProIleProValThrProLysSerProTrpSerMetAspGluA 184
551 ACCTCATGACATGACGCTACGAGGCTGGAATCCCTGAGAACCCCAAGAC 600
184 snLeuMetHisLeuSerTyrGluAlaGlyIleLeuGlnAsnProLysAsn 200
601 CAGGCCCTCCAGGCTCTACACGAAGCCAGACCCAGCCCAAGCCCC 650
201 GlnAlaProGlyLeuTyrThrIlysthnglnAsnProAlaLysAlaPr 217
651 CAACACCCCTGACATTCCTGAGATGAGTCAAAAAAGGGCTCCCTGTA 700
217 oAnThrProAspValLeuGlnIleGluPheLysLysGlyValProValL 234
701 AGGTGACCAAGCTCAAGATGGACACCCAGCCACCTCTGAGAGCTC 750
751 TTCATGTACTGAACGAGTGGGGGCAAGCATGCGTGGCCGATTTGA 800
251 PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValAlaArgLLeas 267
801 CATGTGGAGAACCCGCTCATGTAATGAAGTCCGAGGATCTACGAGA 850
267 rIleValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrGluT 284
851 CCCAGAGGACCATCCTTACCATCCTCATTTAGACATGAGGCTTTC 900
284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAlaPhe 300
901 AACATGACCGGGAAGTCCGCAAAATGAACAAGCCTGGGCTTGAATTT 950
301 ThrMetAspArgGlnValArgLysIleLysGlnIleLeuGlyLeuLysPh 317
951 TGGTGAAGTGGTATACCGGTTACGGCTAGCCCTGAGTGAATTTG 1000
317 eAlaIleLeuValTyrThrIlysthnglnAsnProGluCysGluPheV 334
1001 TCCGCGACATGATCGCAAGTCCAGAGAGCATGAGGAAGGAAGTCCAG 1050
334 aArgHisCysIleAspLysSerGlnGluArgValGluGlyLysValGln 350
1051 GTTCCGGTCCCAAGGGCCAGTGTACATCCTCGCGGAGAGTCCCACT 1100
351 ValSerValPheLysGlyGlnValTyrIleLeuLysArgLysSerProLe 367
1101 GTTCTCTACATGAGAGCTGGTGAAGCATGACGAGGCTGATTTATG 1150
367 uSerLeuTyrAsnGlnGluLeuValSerMetAsnValGlnIlyAspTyrG 384
1151 AGGCAACTGATGCCACCGGTTCAACATCAATTCCTCAGGCTGAG 1200
384 lUpTroLleAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATATCATGCTGTCAGAGCAAGGTCACTGCCAANA 1236
401 GluTyrHisArgLeuLysSerLysValThrAlaLys 412

```

```

seq_name: p1r1:AJMSRS
seq_documentation_block:
  argininosuccinate synthase (EC 6.3.4.5) - mouse
  N:Alternate names: citrulline--aspartate ligase
  C:Species: Mus musculus (house mouse)
  C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Jun-1999
  C:Accession: J00463
  R:Sunh, L.C.; Beaudet, A.L.; O'Brien, W.E.
  A:Title: Molecular characterization of the murine argininosuccinate synthetase locus.
  A:Reference number: J00463, MIMD:91216457
  A:Accession: J00463
  A:Molecule type: mRNA
  A:Residues: 1-412 <SUR>
  A:Cross-references: GB:M31690; NID:g192068; PIDN:AAA37266.1; PID:g309111
  C:Comment: This enzyme catalyzes the formation of argininosuccinate from citrulline a
  C:Genetics:
  A:Gene: ASS
  A:Introns: 35/3; 58/3; 121/3; 142/3; 495/3; 189/2; 199/3; 230/1; 258/2; 280/1; 324/1;
  C:Superfamily: argininosuccinate synthase
  C:Keywords: arginine biosynthesis; homotrimer; ligase; urea cycle
  F:149/133/Binding site: Mg-ATP (Glu, Arg) #status predicted

alignment_scores:
  Quality: 2073.00      Length: 412
  Ratio: 5.093          Gaps: 0
  Percent Similarity: 98.786      Percent Identity: 96.117

alignment_block:
  US-09-775-693-1 x AJMSRS  ..

Align seg 1/1 to: AJMSRS from: 1 to: 412

1  ATGTCCAGAAAGGCTCCGGTGTGCTGACCTACAGTGGCGGCTGACAC 50
1  MetSerLysGlySerValValLeuAlaIlySerGlyLysAspThr 17
51  CTCGTGATCCCTGCTGCTGCTGAGAGAACAGCTTACGATGCTACGCT 100
17  rSerCysIleLeuValThrPheLysGlnGlnIlyTyrAspValIleAlaT 34
101  ATGTGCGCAACATGTGGCCAGAGAGACTTGCAGAGCCAGCAAGAG 150
34  yTLeuAlaAsnIleGlyGlnLysGlnAspPheGluGlnAlaArgLysLys 50
151  GCAGTGAAGCTTGGGGCCAAAGGTGTTCAATTGAGATGTCAGCGCA 200
51  AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerLysG1 67
201  GTTGTGAGAGAGTTCATGCTGCGCGGCGCATCCAGTCCAGGCACTGAT 250
67  uPheValGlnGluPheIleThrProAlaValGlnSerMetAlaLeuTyrG 84
251  AGGACCGTACTCTCGGACACTCTCTTGCAGAGCCCTGCATGCGCCG 300
84  lAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 100
301  AAACAAGTGAATGCGCCAGCGGAGGAGGCGCAAGTATGTGCCACGG 350
101  ArgGlnValGluIleAlaGlnArgGluValAlaLysTyrValSerHisG1 117
351  GCGCCAGGAAGGGGAGACGATGAGTCCGGTTGAGCTGAGCTGCTACT 400
117  yAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuThrCysTyrS 134
401  CAGTGGCCCCCAGATAAAGTCATGCTGCTCCCTGGAGAGATCCGTAATC 450
134  eTLeuAlaProGlnIleLysValIleAlaProTrpArgMetProGluPhe 150
451  TACAAACGGTTCAAGGGCCCAATGAGCTGATGGATCGCAAGCAACA 500

```

```

151  TTTTAAATGPhenylsGlyArgAsnAspLeuMetGluTyrAlaLysGlnH 167
501  CGGATTTCCATCCCGGTCACTCCCAAGAACCGGTGAGATGATGAGA 550
167  sGlyIleProValThrProLysSerProTyrSerMetAspGluA 184
551  ACCATGACATCATGCTACAGAGCTGGAATCTTGAGAGAACCCCAAGAC 600
184  snLeuMetHisIleSerTyrGluAlaGlyIleLeuGlnsnProLysasn 200
601  CAAGCGCTCCAGGTCTCTACAGAGACCCAGAGACCCCAAGACCC 650
201  GlnAlaProProGlyLeuTyrThrTyrGlnAspProAlaLysAlaPr 217
651  CAACACCCCTGACATTCGAGATCGAGTTCAAAAAGGGTCCCTGGA 700
217  oAsnSerProAspValLeuGlnIleGluPheLysGlyAlaProValL 234
701  AGGTGACCAACCTCAAGATGGCACCACCCAGACCTCTTGAGAGCTC 750
234  yValThrAsnIleLysAspGlyThrThrArgThrThrSerLeuGluLeu 250
751  TTTCATCTACCTGACAGAGTCCGCGGCAAGCATGGCGGCGCCGCTAT 800
251  PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValGlyArgIleAs 267
801  CATCGTGGAGAACCGCTTCATTTGAATGATCCCGAGTATCTACAGAGA 850
267  rIleValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrGluT 284
851  CCCGAGCAGACACATCTTACCATGCTGCTTACATGAGATGAGCCCTTC 900
284  hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAlaPhe 300
901  ACCATGACACCGGGAAGTCCGCAAAATCAACAGAGCTGGAGCTTGAAT 950
301  ThrMetAspArgGlnValAlaArgLysIleLysGlnIleLeuGlyLeuLysph 317
951  TGCTGAGTGGTATACCGGTTCAGGCTTACGCTGAGCTGAGTGAATTTG 1000
317  eAlaIleuValTyrThrGlyPheThrPheHisSerProGlnCysGlnPheV 334
1001  TCCGCGCATGTCATGCCAAGTCCAGAGAGAGTGAAGGGAAGAGCGAC 1050
334  aArgHisCysIleGlnLysSerGlnLysArgValGlnGlyLysValGln 350
1051  GTGTCCGCTCAAGGCGCAGGTGTACATCTCGCGCGGAGTCCCGCAT 1100
351  ValSerValPheLysGlyGlnValTyrIleLeuGlyAlaArgLysSerProLe 367
1101  GTCTCTACATGAGAGCTGGTGGAGCATGAGTGGAGGCTGATTATG 1150
367  uSerLeuTyrAsnGlnGlnIleuValSerMetAsnValGlnGlyAspTyrG 384
1151  AGCAATGATGCCACCGGCTTCATCAACATCAATTCCTCAGGCGGAG 1200
384  LurProIleAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201  GAATATCATGCTCCAGAGCAAGTCACTGCCAAA 1236
401  GluTyrHisArgLeuGlnSerLysValThrAlaLys 412

```

seq_name: p1r1:AJBORS

seq_documentation_block:

argininosuccinate synthase (EC 6.3.4.5) - bovine

N.Alternate names: citrulline--aspartate ligase

C.Species: Bos primigenius taurus (cattle)

C.Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 23-Feb-1997

C.Accession: A33986

R.Dennis, J.A.; Healy, P.J.; Beaudet, A.L.; O'Brien, W.E.

Proc. Natl. Acad. Sci. U.S.A. 86, 7947-7951, 1989

A>Title: Molecular definition of bovine argininosuccinate synthase deficiency.
A:Reference number: A33986; MUID:90046714
A:Accession: A33986
A:Molecule type: mRNA
A:Residues: 1-412 <DEF>
A:Cross-references: GB:M26198
C:Comment: This enzyme catalyzes the formation of argininosuccinate from citrulline and aspartate.
C:Superfamily: argininosuccinate synthase
C:Keywords: arginine biosynthesis; homotrimer; ligase; urea cycle
F:149,153/Binding site: Mg-ATP (Glu, Arg) #status predicted

alignment_scores:

Quality	Length	Gaps
2061.00	412	0
Ratio: 5.039		
Percent Similarity: 99.272	Percent Identity: 94.903	

alignment_block:

US-09-775-693-1 x AJBORS

Align seg 1/1 to: AJBORS from: 1 to: 412

```

1  ATGTCAGCAAGAGCTCCGTGTTGCTGCTACAGTGGCGGCTGACAC 50
1  MetSerGlyLysGlySerValValLeuAlaTyrSerGlyLeuAspThr 17
51  CTCTGATCTCTGTTGCTGAGAGCAAGAGCTATGAGCTATGCTCT 100
17  rSerCysIleLeuValTrpLeuLysGlnGlnGlyTyrAspValIleAla 34
101  ATCTGGCCCAACATGGCCGAGAGAGAGAGACTTCAGAGAGAGAGAG 150
34  yrLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGlnAlaAlaArgLys 50
151  GCATGAGAGCTTGGGCGCAAAAGTGTTCATGAGAGAGAGAGAGAG 200
51  AlaLeuLysLeuGlyAlaLysLysValPheIleGlnAspIleSerLys 67
201  GTTGTGAGAGAGTTCATGCTGGCGGCTGACAGTCCAGGCACTGATG 250
67  uPheValGlnIlePheIleThrProAlaIleGlnSerSerAlaLeuTyrG 84
251  AGGACCGCTACCTCTGGGACAGCTCTGTCAGGCGCCGAGAGAGAG 300
84  LysAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 100
301  AAACAGATGGAATGCCCGCAGCGGAGAGGCGCCAGATATGATGCCAG 350
101  LysGlnValIleIleAlaGlnArgGlnGlyAlaLysTyrValSerHisG 117
351  CGGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400
117  yAlaThrGlyLysGlyAsnAspGlnIleArgPheGlnLeuThrCysTyrS 134
401  CACTGGCCCCCAGATAAAGTCAATGCTCCCTGAGAGAGAGAGAGAG 450
134  eArgLeuAlaProGlnIleLysValIleAlaProTyrArgMetProGlu 150
451  TACAACCGGTTCAGAGGCGGCAATGACCTGATGAGAGAGAGAGAG 500
151  TyrAsnArgPheGlnIleLysAsnAspLeuMetGluTyrAlaAsnGlnH 167
501  CGGATTCCTCATCCGCTCCTCCCAAGACCCGCGAGAGAGAGAGAG 550
167  sGlyIleProValProValThrProLysAsnProTyrSerMetAspGlu 184
551  ACCCTATGACATGACGCTACAGAGCTGGAATCTGAGAGAGAGAGAG 600
184  snLeuMetHisIleSerTyrGluAlaGlyIleLeuGlnAsnProLysasn 200
601  CAAGCGCTCCAGGTCTCTACAGAGAGAGAGAGAGAGAGAGAGAGAG 650
201  GlnAlaProProGlyLeuTyrThrLysThrGlnAspProAlaLysAlaPr 217

```


[illegible]

7 AGCAAGAGCTCCGTGGTCTCGGCGCTACAGTGGCGGCTCGACACCTGGT 56
2 SerTysGlyValLysLeuAlaIyrSerGlyGlyLeuAspThrSerVa 18
57 CATCTCGTGGGTGAGAGAACAGGCGTATGACGTATTGCCATTATCTGG 106
18 IleLeuAlaIyrPheLeuAspThrGlnGlyIyrGluValAlaIaIaPheMetA 35
107 CCACATTGGCCAGAACAGACCTTGAGGAAGCCAGGAGAACAGGACACTG 156
35 IaasVaIaIyGlnGlnGlnAspPheAspAlaAlaIySduIuYsAlaLeu 51
157 AAGCTTGGGGGCCAAAAGAGTGTTCATTGAGAGTGTCCAGGAGCTTTGT 206
52 LysIleGlyAlaIcyLysPheValCysValAspCysArgGluAspPheVa 68
207 GGAGGAGTTTCATCTGGCGCGGCACAGTCACAGCGACGTGTGAGAGACC 256
68 LlysAspIleLeuPheProAlaValAlaGlnValAsnAlaValIyrGluAspY 85
257 GCATCTCTCTGGGCACCTCTTCTTCCAGCGCCCTGCATCGCCGCAACAA 306
85 aIlyrIleuLeuIlyrSerLeuAlaIyrProValIleAlaIyAsnAlaGln 101
307 GTGGAATCGCCACGCGGAGGGGGCCAAAGTATGTGTGCCACGGCCAC 356
102 IleAspValAlaIySerIleGlnIcySphenAlaValSerHisGlySerH 118
357 AGGAAAGGGAAACGATCAGTCGCGTTGTGACCTCAGCTCAGTCACTGG 406
118 rGlyLysGlyAsnAspGlnIleArgPheGluLeuSerPheTyrAlaLeu 135
407 CCCCACGATAAAGTCATTGCTCCCTCGAGAGATGCTGAAATCTACAC 456
135 ySProValValIyScySIleThrProIraPheMetProGluPheMetGlu 151
457 CGGTTCAGAGGCGCGCATGACCTGATGAGATGACGCAACAGCACGGAG 506
152 ArgPheAlaIcyArgLysAspLeuAsnAspTyrAlaIaIaGlnIyGlyI 168
507 TCCCATCCGGCTACTCCACAGACCCGTGGAGCATGATGATGAGACCTCA 556
168 eProValAlaGlnIThrLysAlaIySerProIraPheThrAspIuSerGlnA 185
557 TGCAATCAGTACGAGCGTGAATCTGTGAAACCCCAAGACCAAGCAGC 606
185 IaHisIleSerTyrGluAlaGlyIleLeuGlnAspProAspThrThrPro 201
607 CCTCCAGCTCTACACAGAACCCAGACCCAGACCCAGCAAGGCCCCACAC 656
202 ProLysAspMetTyrPylSerIleuAlaValAspProMetAspAlaProAspG 218
657 CCTGCAGTTCCTCCAGATTCAGTTCAAAAAGAGGGTCCCTGTGAAGTGA 706
218 nProGlnAspLeuThrIleAspPheGlnArgGlyLeuProValIySLeuT 235
707 CCAAGCTCAAGATGACACCAACCCACACACCTCC.....TTG 744

235 hTyrThrAspAsnLysThrSerLysGluValSerValThrLysProLeu 251
 745 GAGCTCTCATCTGACCTGAACGAGTCGCGGCAAGCATGGCGTGGCG 794
 252 AspValPheLeuAlaIAserAsnLeuAlaArgAlaAsnGlyValGlyAr 268
 795 TATTGACATCTGGAGAACCGCTTCATTGTGAATGAATCCGAGGATCT 844
 268 gileAspIleValGluAspArgTyrIleAsnLeuLysSerArgGlyCys 285
 845 ACAGACCCCGACAGACGACCATCTTACCATGCTCATTTAGACATCGAG 894
 285 yreGluInAlaProLeuThrValIleArgLysAlaHisValAspLeu 301
 895 GCCTTCACCATGACCGGAGTGGCAAAATCAACAGAGC...CTGGG 941
 302 GlyLeuThrLeuAspLysGluValArgLysLeuLysAspSerPheValTh 318
 942 CTTGAAATTTGCTGAGCTGGTGTATACGGGTTTACGGCTAGCCCTAGT 991
 318 rProAsnTyrSerArgLeuIleTyrAsnGlyPheLeuLeuHisProGlu 335
 992 GTGAATTTGCTCCGACATGTCATCCGCAAGTCGCGAGACGAGTGAAGG 1041
 335 ysgLutTyrIleArgSerMetIleGlnProSerGlnAsnSerValAsnGly 351
 1042 AAAGTCAGAGTGTCCGCTTCACAGGCGAGGTGTACATCCTCGCGCGGA 1091
 352 ThrValArgValArgLeuTyrGlyAsnValIleIleLeuLysArg 368
 1092 GTCCCACTG...TCTCTACATATGAGAGCTGTCGATGATGATGATG 1138
 368 rThrLysThrGluLysLeuTyrAspProThrArgLysSerMetAspGlu 385
 1139 AGGTGATTTAGACCAATGATGATGATGATGATGATGATGATGATGAT 1188
 385 eutHrGlyPheLeuProThrAspThrThrGlyPheIleAlaIleGlnAla 401
 1189 CTCAGGCTGAAGAAATATCATCTGTCACAGACAGATGATGAT 1230
 402 IleArgIleLysLysTyrGly.....GluSerLysLysThr 413
 seq_name: p1r2:D75490

seq_documentation_block:
 argininosuccinate synthase - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: D75490
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Smith, H.O.; Vamathevan, J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
 Science 286, 1571-1577, 1999
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: D75490
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-402 <WHI>
 A:Cross-references: GB:AE001924; GB:AE000513; NID:g6458372; PIDN:AAF10250.1; PID:g645837
 C:Genetics:
 A:Gene: DR0674
 A:Map position: 1
 C:Superfamily: argininosuccinate synthase

alignment_scores:
 Quality: 1034.00 Length: 411
 Ratio: 3.272 Gaps: 6
 Percent Similarity: 76.886 Percent Identity: 50.122
 alignment_block:

US-09-775-693-1 x D75490 ..
 Align seg 1/1 to: D75490 from: 1 to: 402

7 AGCAAGAGCTCCGATGTTCTGCTACAGTGGCGGCTGGACACCTGCTG 56
 2 SerLysGluLysIleValLeuAlaTyrSerGlyGlyLeuAspThrSerI 18
 57 CATCCCTCGTGGCTGAAG...GAACAAGCTATGACGCTATTCCTATC 103
 18 eileLeuLysTyrLeuGlnThrGlnArgAsnTyrAspValValCysPhe 35
 104 TGGCCACATGTCGCGAGAGAGACTTCGAGAGAGCGAGAGAGAGCA 133
 35 hTAlaAspLeuGlyGlnGlyAspGluValGlnGluAlaArgValLysAla 51
 154 CTGAAGCTTGGGCGCAAAAGGTGTTCATTGAGATGTCAGAGGAGTT 203
 52 LeuAsnThrGlyAlaValAlaIleTyrAlaLeuAspLeuArgGluLys 68
 204 TGTGAGAGAGTTCATCTGGCGGCGCATCCAGTCCAGGCTGATGAG 253
 68 eValArgAspTyrValPheProMetMetArgSerSerAlaLeuTyrGlu 85
 254 ACCGCTACCTCTGCGACCTTCTTTCGCGCGGCTGATCGCGCGCA 303
 85 LyrTyrTyrLeuLeuGlyThrSerIleAlaArgProLeuIleAlaLys 101
 304 CAAGTGAATTCGCGCGCGGCGGCGGCGCAATATGTCCTCCAGCGCG 353
 102 MetValGlnIleAlaGlnLysGlnGlyAlaValAlaIleSerHisGlyAl 118
 354 CACAGGAAGGGAGACATCAGTCCGATGTTGAGTCACTGCTCTGAC 403
 118 atHrGlyLysGlyAsnAspGlnValArgPheGlnMetSerAlaTyrAla 135
 404 TGGCGCGCGATTAAGGTCATGCTCCCTGAGAGATGCTTAATTCAC 453
 135 eulysProAspIleValThrValAlaProThrArgAspThrAsp..... 149
 454 AACCGGTTCAAGGCGCGCAATGATGATGATGATGATGATGATGATG 503
 150PheGlnGlyArgAlaAspLeuGluAlaPheAlaArgLysIle 164
 504 GATTCCATTCGCGCTACCTCCCAAGACCGCTGGAGCATGATGAGAAC 553
 164 ylleProValProThrThrLysLysAspProThrPserMetAspAlaAsn 181
 554 TCATGCATCATGATCAGAGGCTGGAATCTGGAACCCCAAGACCAAC 603
 181 etLeuHisIleSerTyrGlnGlyGlyProLeuGlnAspProThrThrGlu 197
 604 GGGCCCTCCAGTCTACACAGAACCCAGACCCCAAGACCCCAAC 653
 198 ProThrThrHisMetPheLysLeuThrValAsnProGluAspAlaPro 214
 654 CACCCCTGACATTCGAGATCGAGATTCAAAAAGGGGCTCCGTGAGG 703
 214 rLisAlaGlnTyrValGlnIleGlnTyrValAsnGlyAspProValSer 231
 704 TGACCAACGTAAGAGTGGACACACCCAGACGACCTCTTGAGCTTC 753
 231 leAsnGly.....GluLysLeuSerProAlaIleLeu 242
 754 ATGTACTGACGAAGTCGCGGCGCAAGCTGGTGGCGCTTTCATCAT 803
 243 ThrLysAlaAsnGlnIleAlaGlyArgHisGlyValGlyValGlyLeu 259
 804 CTGAGAGACCGCTTCATGGAATGAGTCCGAGGTATCTACGAGACCC 853
 259 uValGluAsnArgPheValGlyMetLysSerArgGlyValTyrGluThr 276
 854 CAGCAGGACACATCCTTACCATGCTCATTTAGACATGAGGCTTCACC 903

```

|||||.....:|||||
276 rogllygthrlleuylrhlslalargalalglulserleuthr 292
904 ATGAGCCGGAACTGCGCAAAATCAAGAGCCGTGGATTTGCG 953
293 LeuAspArgGluValLeuHisGlnArgAspAlaLeuGlyProLysTyrAl 309
954 TGAAGCTGTATACCGGTTTACGGCCGTAGCCCTGAGTGTGAA..... 996
309 aglulevaltyrasncllyrhetrrpheelarprogluarglualaleug 326
997 .....TTTGTCCGCACTGCATGCCAAGTCCAGAGCGAGTGGAAAGG 1041
326 lvaltyrphedprhl...ValAlaLysSer.....ValThrGly 338
1042 AAATGCAAGGTGTCCTGCAAGGCGCCAGGTGATACCTGCTCCGCGGGA 1091
339 ThrAlaArgLeuLysLeuTyrLysGlyAsnGlyAlleValAlaGlyArgly 355
1092 GTCCCACTGTCTCTACATGAGAGAGCTGTGAGCATGACGTGCAGG 1141
355 salaglunargserleuylrAspLysAspLeuValSerPheGluAlaGly 372
1142 GTGATTATGACCCAACTGATGCCACCGGTTTCATCAATCAATTCCTTC 1191
372 lYAspTyrAsnGlnHisAspAlaGlyAlaPheIleLysLeuAsnSerLeu 388
1192 AGCGTGAAGGAATATCATCTGCTCCAGAGCAAG 1224
389 ArgMetArgValGlnLysArgValGlnAspLys 399

seq_name: plr2:C83204
seq_documentation_block:
  argininosuccinate synthase PA3525 (imported) - Pseudomonas aeruginosa (strain PA01)
  C:Species: Pseudomonas aeruginosa
  C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
  C:Accession: C83204
  R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br
  adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardly, K.; Lam,
  ; Lory, S.; Olson, M.V.
  A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
  A:Reference number: AB2950; MUID:20437337
  A:Accession: C83204
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-405 <STO>
  A:Cross-references: GB:AE004773; GB:AE004091; NID:9949672; PIDN:AA606913.1; GSPDB:GN001
  A:Experimental source: strain PA01
  C:gene: argG; PA3525
  C:genetics:
  A:gene: argG; PA3525
  C:superfamily: argininosuccinate synthase

alignment_scores:
  Quality: 927.50 Length: 404
  Ratio: 3.011 Gaps: 6
  Percent Similarity: 76.238 Percent Identity: 47.030

alignment_block:
  us-09-775-693-1 x C83204 ..

Align seg 1/1 to: C83204 from: 1 to: 405
1 ATGTCACAGCAAGGCTCCGTGCTTCCATGAGTGGCGGCTGACAC 50
|||||.....:|||||
1 MetaLAspValLysLysValValLeuAlaTyrSerIleGlyLeuAspTh 17
51 CTCGTCATCTCGTGTGCTGAAGAA...CAAGCTATGACGTCAATG 97
|||||.....:|||||
17 rserValIleLeuLysTrpLeuGlnAspThrTyrAsnGlyValVal 34

```

```

98 CCTATCGCCCAACATTTGCGCAGAGAGACTTCGAGAAAGCCAGAG 147
34 hrPheThrAlaAspLeuGlnGlyGlnValAlaProLAlaArgAla 50
148 AAGCACTGAAGCTTGGGCGCAAAAGGTTCATTTAGAGATGATGAG 197
51 LysAlaArgAlaMetGluValLysGluIleTyrLLeAspAspLeuArg 67
198 GAGTTTGTGAGAGTCACTGCGCGGCGCATTCGATCCAGCCGACT 247
67 uGluPheValArgAspPheValTyrProMetPheArgAlaAsnThrVal 84
248 ATGAGCAAGCTATCTTCGCGCACTCTTTCGAGCGCCGTGATGCG 297
84 yrgluGluGluTyrLeuLeuGlyThrSerIleAlaArgProLeuIleAla 100
298 CGCAACAAAGTGAATGCGCCAGCGGAGCGGCAAGTATGTGTGCA 347
101 LysArgLeuIleGluIleAlaAsnGluThrGlyAlaAspAlaIleSer 117
348 CGGCGCCACAGAAAGGGAAGCATCAGTCCGCTTGGAGCTGACTGCT 397
117 SGlyAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuGlyAla 134
398 ACTCATGCGCCCGCAATAAAGTCAATGCTGCTCCGAGATGAGTGA 447
134 yAlaLeuLysProGluValLysValIleAlaProThrPArgLutPasp 150
448 TTCATCAACCGGTTCAAGCGCGCATGACCTGATGAGTACGCAAGCA 497
151 LeuLeuSerArgLys.....LeuMetAspTyrAlaGluLys 163
498 ACAGCGGATTCCTATCCG.....GTCATCCCAAGAACCCGCGAGCA 541
163 shISgIleProlleGluArgHisGlyLysLysIleSerProLysIle 180
542 TGGATGAGAACTCATGACATCAGTACGAGGCTGGAATCCTGAGAAC 591
180 etAspLAsnLeuLeuHisIleSerTyrGluGlyGlyValLeuGlnAsp 196
592 CCAGAACCAAGCGGCTCCAGCTGCTACAGAGAACCCAGAGCCAGC 641
197 ThrTrpThrGluHisGluLysAspMetTrpLysTrpThrAlaSerPro 213
642 CAAGCCCCCAACACCCGATTCGATTCGAGTCAAGTCAAAAAGGG 691
213 uAsnAlaProAspThrProThrTyrIleGluLeuThrTyrArgLysGly 230
692 TCCCTGTGAAGTGAACCAAGTCAAGATGAGCAACCCACAGACTCC 741
230 splleValAlaIle.....AspGly...LysAspMetThrPro 241
742 TTGGAGCTTTCATGTACCTGAGCAAGTCCGCGGACAGCATGCGTGG 791
242 AlagluValLeuThrGluLeuAsnArgValGlyGlyIleAsnGlyIle 258
792 CCGTATGACATCGTGGAGAACCGCTTCATTTGATGAAGTCCAGAGTA 841
258 yArgLeuAspIleValGlnAsnArgTyrValAlaGlyMetLysSerArg 275
842 TCTAGAGACCCAGCAGGACCATCTTCATTTGATGAAGTCCAGAGTA 891
275 ystyGluThrProGlyGlyThrIleMetLeuLysAlaHisArgAlaIle 291
892 GAGGCTTCACCATGACCGGAGAGTCCGCAAAATCAAGAGCCGCTGG 941
292 GluSerIleThrLeuAspArgGluValAlaHisLeuLysAspLLeuLeu 308
942 CTGGAATTTGCTGAGCTGTGTATACCGTTTACGCGCTAGCCGATG 991
308 TrProLysTyrAlaSerLeuIleTyrThrGlyTyrTrpPserProGlu 325
992 GTGATTTTGCGGCACTGATGCCAAGTCCAGAGCGAGTGAAGG 1041

```

```

325 rgleuMetLeuGlnGlnMetIleAspAlaSerGlnValAsnValAsnGly 341
1042 AAAGTCGAGGTCGTCCTGACGAGGCGGACGTCATCTCGGCGGGA 1091
342 ValValAlaArgLeuLeuSerGlyAsnValValValValGlyArg 358
1092 GTCCCGCCTCTCTCTACATGAGGAGCTGTGAGCATGACGTCGACG 1141
358 sSerAspAspSerLeuPheAspAlaSerGlnValAlaThrPheGlnGluAspG 375
1142 GTGAT...TATGAGCCACTGATGACCGGCTTCAATCAATTC 1188
375 TylGlyAlaTyrAsnGlnAlaAspAlaAlaGlyPheIleLeuAsnAla 391
1189 CTCAGGCTGAAG 1200
392 LeuArgMetArg 395

```

seq_name: p1r2:C81415

```

seq_documentation_block:
  argininosuccinate synthase (EC 6.3.4.5) Cj0665c [Imported] - Campylobacter jejuni (strain
  C:Species: Campylobacter jejuni
  C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
  C:Accession: C81415
  R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling
  Nature 403, 665-668, 2000
  A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf
  A:Reference number: AB1250; M01D:20150912
  A:Accession: C81415
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-406 <PAR>
  A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CA875297.1; PID:9696812
  A:Experimental source: serotype O2, strain NCTC 11168
  C:Genetics:
  A:Gene: argG; Cj0665c
  C:Superfamily: argininosuccinate synthase
  C:Keywords: ligase

```

```

alignment_scores:
  Quality: 892.00      Length: 397
  Ratio: 3.014        Gaps: 6
  Percent Similarity: 74.559      Percent Identity: 47.355

```

alignment_block:

US-09-775-693-1 x C81415 ..

Align seg 1/1 to: C81415 from: 1 to: 406

```

19 GTGGTTCTGGCTACAGTGGCGGCGGACCTCTGTCATCTGCTGTG 68
   |||||
8 ValValLeuAlaTyrSerIleGlyLeuAspThrSerIleLeuLeuTyr 24
   |||||
69 GCTGAAG...GAACAAGCTATGACCTATGCTATCGGCAACATG 115
   |||||
24 pleuGlnAspGluTyrAsnGlyValValAlaThrPheThrAlaAspIleG 41
   |||||
116 GCCAGAGGAGAGACTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 165
   |||||
41 TylGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 57
   |||||
166 GCCAAAAG...GTGTCATTGAGAGTGTGACGAGGAGGAGGAGGAGGAG 209
   |||||
58 TleTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 74
   |||||
210 GGAGTTCATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 259
   |||||
74 saspyrValPhePrometPheArgAlaAsnAlaIleTyrGlnGlnGln 91

```

```

260 ACCTCTGGGACCTCTCTGTCGAGGCGGCGGCGGCGGCGGCGGCGGCA 309
   |||||
91 YrleuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 107
   |||||
310 GAAATGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 359
108 GlnIleAlaLeuGlnThrGlyAlaAspAlaValSerHisGlyAlaThrG 124
   |||||
360 AAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 409
124 YrlysglyAsnAspGlnValAlaThrPheGlnGlnGlnGlnGlnGln 459
   |||||
410 CCCAGATAAAGGTCAATGCTCTGCTGAGAGAGGAGGAGGAGGAGGAG 499
141 roAspLeuTyrIleIleAlaProTyrGlnGlnGlnGlnGlnGlnGlnGln 509
   |||||
460 TTCAGGCGGCGGAGTACCTGATGAGTACGCAAGGAGGAGGAGGAGG 509
158 GlnTyr...LeuLeuAlaTyrAlaGlnGlnGlnGlnGlnGlnGlnGln 510
   |||||
510 CATCCCGGTCACTCC...AAGAACCGGTGAGGAGGAGGAGGAGGAGG 556
170 rleSerTyrLysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 187
   |||||
557 TGCACATGACGTCAGAGGCTGGAATCTGTGAGAACCCCAAGGAGGAG 606
187 euhIleSerTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 606
   |||||
607 COTCAGGCTCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 656
204 GlnGlnAspMetTyrAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
   |||||
657 CCCGACATTTCTGAGATGAGTTCGAAAGGAGGAGGAGGAGGAGGAGG 706
220 usGlnIleIleIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 237
   |||||
707 CCAACGTCACAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 756
237 snGlyGlnTyr...LeuSerProAlaGlnGlnGlnGlnGlnGlnGlnGln 248
   |||||
757 TACCTGACGAGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 806
249 LysLeuAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 265
   |||||
807 GGAGAACCGGCTGATGAGTGAATGAGTCCGAGGAGTTCAGGAGGAGG 856
265 IgluAsnArgTyrValGlyMetLysSerArgGlyGlyGlnGlnGlnGln 282
   |||||
857 CAGGACCATCTTTCATGCTCTGATGATGATGATGATGATGATGATG 906
282 TylGlyThrIleLeuLeuLysAlaHisAlaGlnGlnGlnGlnGlnGln 298
   |||||
907 GACCGGAGAGTGGCAAAATCAACAAGGAGGAGGAGGAGGAGGAGGAG 956
299 AspArgGlnAlaAlaHisLeuLysAspGlnGlnGlnGlnGlnGlnGln 315
   |||||
957 GCTGTGTATTCGCTTACGCGGCTTACGCGGCTTACGCGGCTTACGCG 1006
315 rleuIleTyrAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 332
   |||||
1007 ACTGCATGCCAGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
332 IalLeuIleAlaAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 348
   |||||
1057 GTCCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1103
349 LeuTyrLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 365
   |||||
1104 TCTTACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1153
365 rleuPheAsnAlaAlaTyrCysThrPheGlnGlnGlnGlnGlnGlnGln 382
   |||||
1154 CAACGTAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1194

```

382 InLysAspAlaIaGlyPheIleLysLeuAsnAlaLeuArg 395
 seq_name: p1r2:C84048

seq_documentation_block:
 argininosuccinate synthase (citrulline-aspurate ligase) argG [imported] - Bacillus halodurans
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000
 C:Accession: C84048
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650, MUID:20263314
 A:Accession: C84048
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-409 <STO>
 A:Cross-references: GB:AP001518; GB:BA000004; NID:q10175792; PIDN:BA006906.1; GSPDB:GN0C
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: argG
 C:Superfamily: argininosuccinate synthase

alignment_scores:
 Quality: 865.00 Length: 410
 Ratio: 2.912 Gaps: 3
 Percent Similarity: 72.439 Percent Identity: 44.878

alignment_block:
 US-09-775-693-1 x C84048 ..

Align seg 1/1 to: C84048 from: 1 to: 409

```

7  ACCAAGAGCTCCGCTGCTTCTGCTACAGTGGCGGCTGACACCCGCTG 56
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2  SerLysLysValValLeuAlaTyrSerLysLysLysLysLysLysVal 18
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
57 CATCTCGTGTGCTGAAGAACAGAGCTATGACGTCATTCCTCATCTG 106
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
18  lAlaIleLysTrpLeuSerAspLysGlyTyrAspValIleAlaValGly 35
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
107 CCACATTTGGCCAGAGAACAGCTCGAGAGAACCCAGAGAACGACCTG 156
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
35  euAspValGlyGlyGlyLysAspLeuGluPheValLysGlyLysAlaLeu 51
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
157 AAGCTTGGGGCCMAAAGGTGTTTCATGAGATGTCACGAGGAGTTTGT 206
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
52  LysValGlyAlaIleGluSerTyrThrIleAspAlaLysLysGluPheAl 68
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
207 GAGAGATTCATCTGCGCCGCGCATCGACGCGCAGGCTATGAGACG 256
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
68  aGluGluPheValLeuProAlaLeuGlnAlaHisAlaLeuTyrGluGln 85
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
257 GCTACTCTCTGGGACCTCTTTCAGGAGGCTGATGCGCCGCAACAA 306
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
85  ySyrProLeuValSerAlaLeuSerArgProLeuIleSerLysLysLeu 101
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
307 GTGGAATCGCCCGAGCGGAGCGGCGCAAGTATGTCCACAGGCGCAC 356
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
102 ValGluIleAlaGluGlnThrGlyAlaGlnAlaValAlaHisGlySer 118
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
357 AGAAGAGGAGAGATCAGATCGGTTGAGTCACTGCTACTCTGCTG 406
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
118  rGlyLysGlyAsnAspGlnValArgPheGluValSerIleGlnAlaLeu 135
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
407 CCCCCAGATAAAGTCACTGCTCCCTCGAGGATGCTGATTCATCAAC 456
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
135  snProAsnLeuGlnValLeuAlaProValArgGluTrrPalatP..... 149
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
457 CGGTTCAAGGCGGCAATGACCTGATGAGTACCAAGAACACGCGGAT 506
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

150  .....SerArgAspGluGluIleGluTyrAlaLysLysAsnAsnI 163
507  TCCCATCCCGGTCACCTCCCAAGAACCCGTCGAGCATGATGAGAACCTCA 556
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
163  eProIleProIleAspLeuAspAsnProTyrSerValAspGlnAsnLeu 180
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
557  TGCACATCAGCTGAGAGGCTGGAATCTGGAGAACCCCAAGAACCAAGCG 606
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
180  rPglYArgSerAsnGluCysGlyIleLeuGluAspProTrrPalatThrPro 196
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
607  CCTCCAGGCTCTTACACAGAACCCGAGACCCAGCAACCCCAACAC 656
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
197  ProGluIleValAlaTyrGluLeuThrValAlaIleGluAspAlaProAspG 213
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
657  CCCGACATTCCTGAGATCGAGTTCAAAAAGGGCTCCCTGGAAGGTGA 706
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
213  nProGluIleValGluIleGluPheGluLysGlyIleProValThrLeu 229
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
707  CCACGTCAGAGATGGACACCCACCAACACCTCTGAGAGCTTTCATG 756
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
230  ..AsnGlyLysSerTyrProValHis.....GluLeuIleLeu 241
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
757  TACCTGACGAGAGTGGCGGCAAGCATGCGTGGCCGTATTGACATGCT 806
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
242  GluLeuAsnGlnIleAlaGlyLysHisGlyValGlyArgIleAspHisVal 258
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
807  GGAGAACCGCTTCATTTGATGAATGATCCGAGATCTACAGACCCAG 856
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
258  lGluAsnArgLeuValGlyIleLysSerArgGluValTyrGluCysProG 275
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
857  CAGGACCATCTTACCATGCTGCTTACATGACATGAGGCCCTTCACCATG 906
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
275  LysAlaMetThrLeuIleLysAlaHisLysGluLeuGluAspLeuThrLeu 291
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
907  GACCGGAAAGTGGCAAAATCAACACAGCCCTGGGCTTGAATTTGCTGA 956
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
292  ThrLysGluValAlaHisPheLysProValValGluLysLysIleAlaG 308
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
957  GCTGGGTATACCGGTTTACGCGCTAGCCCTGAGTGAATTTGCTGCC 1006
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
308  uLeuIleTyrGluGlyLeuTrrPheSerProLeuGlnProAlaLeuSerA 325
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1007  ACTGCATCGCCAAAGTCCAGAGAGCGAGTGGAGGAAGGAACTGAGCTGC 1056
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
325  lApeLeuLysGluTrrGlnSerThrValThrGlyValValAlaValLys 341
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1057  GTCCTCAAGGCGCGGTGATCTCTGCGCGGAGTCCCATCTCTCT 1106
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
342  LeuPheLysGlyHisAlaIleValGluGlyArgLysSerGluTyrSerLe 358
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1107  CTACATGAGAGCTGTGAGCATGAACGTGCGAGGCTGATTTAGACCA 1156
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
358  uTyrAsnGluLysLeuAlaThrTyrThrProAspAspGluPheAspHisA 375
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1157  CTGATCGCACCGGCTGATCAACATCAATTCCTCAGGCTGAAGCATAT 1206
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
375  snAlaIleValGlyPheIleSerLeuTrrPglYLeuProThrLysValTyr 391
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1207  CATGCTTCCAGAGCAAGTCACTGCTCCAAA 1236
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
392  SerMetValAsnLysGluMetLysGluLys 401
  ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_documentation_block:
 argininosuccinate synthase argG - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: B69589
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azvego, V.; Ber
 C:Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997

```

alignment_scores:      Quality: 818.00      Length: 400
                       Ratio: 2.773      Gaps: 3
Percent Similarity:    73.750      Percent Identity: 42.750

alignment_block:
US-09-775-693-1 x B6589      ..

Align seg 1/1      to: B6589      from: 1      to: 403

```

[illegible]

```

501 CGGGATTCCATCCCGGTCTACTCCCAAGAACCCGTGAGCATGATGAGA 550
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162 gellilleProleProleInsLeuAspSerProTyrSerIleAspIln 179
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
551 ACCTCATGCACATGACGTACGAGCGCTGGATTCCTGGAGAACCCCAAGAAC 600
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 snleuTrpGlyArgIlaInsInLucyselIleuInLucsProTyrIla 195
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
601 CAAGGGCCCTCAGGCTCTCTACACAGAACCCAGACCCAGCAAGACCC 650
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 AlaProProGluGlyAlaTyrAspLeuThrIlaProLeuGlnLysThrPr 212
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
651 CAACACCCCTGACATTCGAGATGCGATGCTCAAAAAGGGGCTCCGTGGA 700
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
212 oAspThrProGluValIleGluIleAlaIleGluGlnGlnGlyAlaProVal 229
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 AGGAGACCAACGTCAAGATGGGACACCACCCACCACTCTTGAGAGCTC 750
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 erIle.....AspGlyValSerTyrSerLeuSer...GluLeu 240
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
751 TTCATGTACCTGAACGAAGTCGCGGGCAAGCAAGTCGCGGGCCGATTTGA 800
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
241 IleLeuLysLeuInsInLucsIleAlaGlyAlaIleGlnGlyAlaGlyArg 257
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
801 CATCGTGGAGAACCGGCTTCATGTGATGATGAAGTCCCGAGTATCTACAGA 850
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
257 rHisValIleGlnAsnArgLeuValGlyIleLysSerArgGluValIleArg 274
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
851 CCCAGACAGCACCATTCCTTACCATGCTCTTAAAGATCGAGAGCGGCTTC 900
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 yspProGlyAlaMetThrIleuIleLysAlaHisLysGluLeuGlnAspLeu 290
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
901 ACCATGGACCGGAAAGTCGCAAAATCAACACACAGCGCTGGCTGAATT 950
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 ThrLeuValLysGluValAlaIleHisPheLysProIleIleGluInLysMe 307
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
951 TGCAGAGCTGCTGATACCGGTTTACCGGCTACCGCTGATGTGAATTG 1000
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
307 LserIleIleIleLysInsnGlnLysLeuThrPheSerProLeuLysAspAla 324
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1001 TCCGCCACTGCATGCGCAAGTCCCGAGAGGAGTGGACAGGAAGTGCAG 1050
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 euHisAlaPheLeuLysGluThrGlnLysHisValIleThrIleValArg 340
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1051 GTGCGCCCTCCTCAAGGCCACGCTGATACATCTCGCGCGGAGATCCCACT 1100
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 ValLysLeuPheLysGlnHisAlaIleValGlnGlyArgLysSerGlnTyr 357
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1101 GTCTCTCTACATGAGAGCTGGTGGAGCATGGAACGTGACAGGTGATTATG 1150
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
357 rSerLeuLysAspGluLysLeuAlaIleThrLysThrLysAspAlaPhe 374
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1151 AGCCAACTGATGCCACCGGTTTCATCAACATCAATTCCTCAGGCTGAG 1200
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
374 sPHisHisAlaIleGlnPheIleGlnLeuLeuTrpGlyLeuProThrLys 390

```

```

seq_documentation_block:
  argininosuccinate synthase (EC 6.3.4.5) - Arabidopsis thaliana
  N:Alternate names: protein F617.40
  C:Species: Arabidopsis thaliana (mouse-ear cress)
  C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
  C:Accession: T06667
  R:Bevan, M.; Terry, N.; Ardies, W.; Buysaert, C.; Dasseville, R.; De Clerck, R.;
  Chelst, J.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
  submitted to the Protein Sequence database, April 1999
  A:Reference number: Z15792
  A:Accession: T06667
  A:Molecule type: DNA
  A:Residues: 1-498 <BEV>
  A:Cross-references: EMBL:AL049657; GSPDB:GN00062; ATSP:F617.40
  A:Experimental source: cultivar Columbia; BAC clone F617

```



```

alignment_scores:      Length: 396
                        Quality: 807.50
                        Ratio: 2.710
                        Gaps: 5
Percent Similarity: 75.253      Percent Identity: 43.683

alignment block:
US-09-775-693-1 x E84935      ..

Align seg 1/1      to: E84935      from: 1      to: 403

```

```

19 GTGTTCTGCTACAGTGGCGGCTGACACCTGTCGATCTCTGTG 68
9 ValValLeuAlaTyrSerGlyLeuAspThrSerAlaIleIleProtr 25
69 GCTGAAGGAGCA... GCGTATGAGCTCATTTGCTGCGCAACATG 115
25 pLeuLysGluAsnTyrAsnPhgGluValAlaIleAlaIleAla 42
116 GCCAG... AAGGAGACTTCGAGAGGAGGAGGAGGAGGAGGAG 162
42 LysIleSerLysLysAspLeuAsnGlyIleGluLysLysSerLeu 58
163 GGGGCGCAAAAGGCTTATTGAGATGTCAGCAGGAGGAGGAGGAG 212
59 GlyAlaSerSerCysHisValPheAspLeuLysGluIleLeu 75
213 GTTCACTGCGCGGCGGCTGACGAGGAGGAGGAGGAGGAGGAG 262
75 nTyrValTyrProValLeuLysThrGlyAlaLeuTyrGlySer 92
263 TCCGCGGAGCTCTCTGCGAGGCGGCTGATCGCGGCGGCGGCA 312
92 euleGlyThrAlaMetAlaArgProIleAlaLysGlnValGlu 108
313 ATGCGCCGAGCGGAGGCGGCAAGTATGTCGCCAGGCGGCGGCA 362
109 LeuAlaLeuAsnIleGlyAlaAsnSerLeuCysHisGlyAlaThr 125
363 GGGGAGGATCAGGCTCGGTTGAGCTGAGCTGACTACAGGCGGCG 412
125 sGlyAsnAspGlnValArgPheGluMetAlaTyrAlaIleAla 142
413 AGATTAAGTCAATGCTCCCTGAGAGATGCTGATTCACACCGGTC 462
142 snLeuAsnValIleAlaProtrParg... GluTrp... AspLeu 154
463 AAGGCGGCAATGAGCTGATGAGTACGCAAGGAGGAGGAGGAG 512
155 AsnSerArgIleSerLeuLysLysLysLysLysLysLysLys 171
513 CCGGCTACCTCCCAAGACCGGTGAGAGATGAGTACGATGAGCA 562
171 rThrAlaThrLeuGluLysIleTyrSerLysAspGluAsnSerTrp 188
563 TCAGCTACGAGCTGAGATCTGAGAGACCCCAAGGAGGAGGAG 612
188 IeSerThrGluGlyLeuLeuLeuAsnProTrpAsnGlnSerAsn 204
613 GGTCTACAGAGACCCGAGACCCGAGGAGGAGGAGGAGGAGGAG 662
205 AspCysTrpSerTrpThrValAsnProGluAspAlaProGlu 221
663 CATTCGAGATCGAGTCAAAAAGGAGGCTGAGAGTGCACGAG 712
221 uTyrValSerLeuGlnLeuLysGlyValValSerValAsnAsn 238
713 TCAAGATGCGACCCACGAGACCTCTGAGCTCTTATGATGCTG 762
238 LnyLys... LeuAsnProLeuLysCysValGluGlnLeu 249
763 AAGGAGTGGCGGCGGAGATGGCTGAGGCGGCTATGACATGAG 812
250 AsnSerLeuGlyAlaLysHisGlyIleGlyArgIleAspIleLeu 266
813 CCGCTCATTTGAGATGAGTCCGAGGATCTACGAGACCCGAGCA 862
266 nArgLeuIleGlyMetLysSerArgGlyCysTyrGluThrProGly 283
863 CCATCCTTACACGCTCATTTAGACATGAGGCTTCACCATGAGCG 912
283 hTleIleuethrAlaIleLysAlaIleGluGlnLeuValLeuAsp 299

```

```

913 GATGCGCGCAAAATCAAGAGCGCTGCTGAATTTGCTGAGCTG 962
300 GluSerPheThrTrpArgGluLysIleGlyLeuGluMetSerSerLe 316
963 GTATACGCGTTTACGGCTTACCGCTGAGTGAATTTGCGGCACT 1012
316 LTyAspGlyArgTrpPheSerProIleArgLysSerLeuGlnAla 333
1013 TCGCGAAGTCCGAGAGCGAGTGAAGGAGGAGGAGGAGGAG 1062
333 LaAspSerLeuSerLeuGluIleThrGlyValIleLeuLysLeu 349
1063 AAGGCGGAGTCTACATCTCGCGGAGTCCGCTGCTCTCTCA 1112
350 LysGlySerValThrAlaValGlnLysSerProAsnSerLeuTyr 366
1113 TGAAGAGCTGAGATGAGTCAAGTCAATTCCTCAGGCTGAG 1162
366 rGluGluTyrAlaThrPheGlyGluAspLysValTyrLysGln 383
1163 CCACGCGTTCATCAACATCAATTCCTCAGGCTGAG 1200
383 LaAspGlyPheIleArgLeuPheSerLeuSerLys 395

```

seq_name: p1r2:B70398

seq_documentation_block:

```

argInnosuccinate synthase - Aquifex aeolicus
C/Species: Aquifex aeolicus
C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A/Reference number: A70300; MUID:9619666
A/Accession: B70398
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-401 <ADP>
A/Cross-references: GB:AE000725; NID:g2983598; PIDN:AAC07170.1; PID:g2983600; GB:AE00
A/Experimental source: strain VF5
C/Genetics:
A/Gene: argG
C/Superfamily: argInnosuccinate synthase

```

alignment_scores:

```

Quality: 800.50 Length: 404
Ratio: 2.741 Gaps: 3
Percent Similarity: 72.277 Percent Identity: 41.832

```

alignment_block:

```

US-09-775-693-1 x B70398
Align seg 1/1 to: B70398 from: 1 to: 401

```

```

10 AAGAGCTCGTGTGCTGAGCTGAGGCGGCTGAGACCTGTCGAT 59
3 LysLysArgValIleLeuAlaTyrSerGlyGlyLeuAspThrSerIle 19
60 CCGTGTGCTGAGTGAAGAGACGCTATGAGCTGATTCCTATGCGCA 109
19 eValArgTrpLeuThrGluLysGlyTyrGluValIleThrLysThrAla 36
110 ACATTTGGCCAGAGAGACTTCGAGGAAAGGAGGAGGAGGAGGAG 159
36 sPValGlyGlnGlyGluLysSerGluIleProGluLysAlaArgArg 52
160 CTGAGGCGCAAAAGGCTTCATTGAGATGTCAGAGGAGGAGGAGT 209
53 AlagGlyAlaIleGluAlaIleValGluAspLeuLysGluThrPheAla 69

```


Wed Feb 13 07:36:06 2002

us-09-775-693-1.rpt

Page 16

CC PATHWAY.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- DISEASE: DEFECTS IN ASS ARE THE CAUSE OF CITRULLINEMIA, AN
 CC AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SEVERE VOMITING
 CC SPELLS AND MENTAL RETARDATION.
 CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X01630; CAA25771.1; -.
 DR EMBL; L00084; AAS1783.1; -.
 DR EMBL; L00079; AAS1783.1; JOINED.
 DR EMBL; L00080; AAS1783.1; JOINED.
 DR EMBL; L00081; AAS1783.1; JOINED.
 DR EMBL; L00082; AAS1783.1; JOINED.
 DR EMBL; L00083; AAS1783.1; JOINED.
 DR PIR; A01195; AJHURS.
 DR MIM; 603470; -.
 DR MIM; 215700; -.
 DR InterPro; IPR001518; Arginosuc_synth.
 DR Pfam; PF00764; Arginosuc_synth.
 DR PROSITE; PS003544; Arginosuc_synth; 1.
 DR PROSITE; PS00564; ARGININOSUCCIN_SYN_1; 1.
 DR PROSITE; PS00565; ARGININOSUCCIN_SYN_2; 1.
 KW Arginine biosynthesis; urea cycle; Ligase; ATP-binding;
 KW disease mutation.
 FT NP_BIND 115 123
 FT VARIANT 14 14
 FT VARIANT 18 18
 FT VARIANT 86 86
 FT VARIANT 118 118
 FT VARIANT 157 157
 FT VARIANT 180 180
 FT VARIANT 192 192
 FT VARIANT 272 272
 FT VARIANT 280 280
 FT VARIANT 304 304
 FT VARIANT 324 324
 FT VARIANT 363 363
 FT VARIANT 363 363
 FT VARIANT 390 390
 FT SEQUENCE 412 AA; 46426 MW; D9083E0655D0716 CRC64;

alignment_scores:
 Quality: 2144.00 Length: 412
 Ratio: 5.204 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-775-693-1 x ASSY_HUMAN ..
 Align seg 1/1 to: ASSY_HUMAN from: 1 to: 412

1 ATGTCGAGCAAGAGCGTCGCTGTCGAGCTGACGTGCGCGCTGAGAC 50
 1 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspThr 17
 51 CTCGTGACCTCTCGTGTGGCTGAAGAACCAAGGCTATGACGATCGCT 100
 17 SerCysIleLeuValTyrPheLysGlyGlnGlyTyrAspValIleAlaT 34
 101 ATCTGGCCCAACATTTGGCCAGGAAGAACTTCGAGAGCCAGGAAGAG 150
 34 TyrLeuAlaAsnIleGlyGlnLysGlyAspPheGlyGlnAlaArgLys 50
 151 GCATGAAAGCTTGGGGCCAAAAGGTGTTCATTGAGATGTCCACAGGGA 200
 51 AlaLeuLysLeuGlyAlaLysLysValPheIleGlnAspValSerAla 67
 201 GTTTGTGGAGAGATTCATCTGGCGGCGCATCTCCAGTCCAGCGCTGATG 250
 67 PheValGlnGluPheIleTyrProAlaIleGlnSerSerAlaLeuTyr 84
 251 AGAACCGCTACCTCTGCGACACTCTCTGCGCAGCGCTGATGCGCGCG 300
 84 LysAspArgTyrLeuLeuGlnGlyThrSerLeuAlaArgProCysIleAla 100
 301 AACACAAGTGGAAATCCGCCAGCGGAGGGGCCAAGTATGTGCCACGG 350
 101 LysGlnValGlnIleAlaGlnArgGlnGlyAlaLysTyrValSerHis 117
 351 CGCCACAGGAAGAGGGAGCATCGATCGCTGCTGACCTGACCTGACTACT 400
 117 ValaThrGlyLysGlyLysAspGlnValArgPheGlyLeuSerCysTyr 134
 401 CACTGGCCCCCAGATAAAGTCATTGCTCCCTCGAGAGATGCTGAATTC 450
 134 GluLeuAlaProGlnIleLysValIleAlaProTyrPheMetProGluPhe 150
 451 TACAACGGGTTCAGAGGCCCAATGACCTGATGAGTACGCAAGAACA 500
 151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGlyTyrAlaLysGln 167
 501 CGGAGATTCCTCCGCTGCTACAGCAAGACCCAGGACCCGCAAGAGATG 550
 167 SgLYIleProIleProValThrProLysAsnProTyrPseMetAspGlu 184
 551 ACCTCATGACATCAGCTACGAGAGGCTGAATCCTGGAGAACCCCAAGAC 600
 184 SLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLysAsn 200
 601 CAAGCCCTCCAGGCTCTACAGCAAGACCCAGGACCCCAAGAGCCCG 650
 201 GlnAlaProProGlyLeuTyrThrLysTyrGlnAspProAlaLysAla 217
 651 CAACACCCCTGACATTCGAGATGAGTTCAAAAGGGGCTCCCTGGA 700
 217 OAsnThrProAspIleLeuGlnIleGlnPheLysLysGlyAlaProVal 234
 701 AGGTGACCAAGCTCAAGATGCGACACCCAGCAGACCTCTTGGAGGCT 750
 234 yValaThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGly 250
 751 TTCATGTACCTGACAGAGTCCGGGCAAGCATGGGCGGCGCGATGGA 800
 251 PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValGlyArgIleAs 267
 801 CATCGTGAGAAACGCTTCATTGGAATGAAGTCCGAGGTATTCAGAGA 850
 267 PLeuValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrGlu 284
 851 CCCAGCAGGACCATCTTACCATGCTGATTTAGACATCGAGGCTTC 900
 284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGlnAlaPhe 300

Wed Feb 13 07:36:07 2002

us-09-775-693-1.15P

901 ACCATGACCGGAGTCCGCAAAATCAACAAGGCGTGGCTGAAAT 950
|||||
301 TTTTAAATGATGATGATGATGATGATGATGATGATGATGATG 317
951 TGGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1000
|||||
317 ealagluuvaltyrthrglyleuargproserproglucysgluphe 334
1001 TCGCCACTGATCCGCAAGTCCAGAGGAGGAGGAGGAGGAGG 1050
334 aalghgscysilealalyserserglualargvalgluglylval 350
1051 GTGTCCCTCCCAAGGCGGAGGATGATGATGATGATGATGATG 1100
351 valservalleuylsglylvaltyrileuylarggluserprole 367
1101 GTCTCTTACATGAGAGAGTGGTGGATGATGATGATGATGAT 1150
367 userleuylrasmngluuvalsermetasnvalinglylasptryg 384
1151 AGCAACATGATCCACCGGTTTCATCAATCAATCAATCAATCA 1200
384 luprothraspalathrglypheileasnileasnileuargleuyls 400
1201 GAATATCATGCTCTCCAGAGGAGGATGATGATGATGATGAT 1236
401 glutyrthrasgluinserylvalthralyls 412

seq_name: SwissProt_39:ASSY_RAT

seq_documentation_block: STRAND: PRT: 412 AA.

AC P09034; 1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, last sequence update)
DT 15-JUL-1999 (Rel. 38, last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
LIGASE).
GN ASS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney; PubMed=3174461;
RX MEDLINE=89016648; Pubmed=3174461;
RA Such L.C., Morris S.M., O'Brien W.E., Beaudet A.L.;
"Nucleotide sequence of the cDNA encoding the rat argininosuccinate
synthetase";
RL Nucleic Acids Res. 16:9352-9352(1988).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC -1- PROPHOSPHATE + L-ARGININOSUCCINATE
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC -1- PATHWAY: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X12459; CAA30999.1; -
DR EMBL: M36708; AAA40771.1; -
DR PIR: S01440; AJRTS.
DR InterPro: IPR001518; Argininosuc_synth.
DR Pfam: PF00764; Argininosuc_synth_1.
DR ProDom: PD003544; Argininosuc_synth_1.
DR PROSITE: PS00564; ARGININOSUCCIN_SYN_1; 1.

DR PROSITE: PS00564; ARGININOSUCCIN_SYN_2; 1.
KW Arginine biosynthesis; Urea cycle; ligase; ATP-binding.
FT NP_BIND 115 123 ATP (POTENTIAL).
SQ SEQUENCE 412 AA; 46496 MW; CCA80906F5A3E93D CRC64;

alignment_scores: Length: 412
Quality: 2086.00 Gaps: 0
Ratio: 5.113
Percent Similarity: 99.029 Percent Identity: 96.845

alignment_block:
US-09-775-693-1 x ASSY_RAT

Align seg 1/1 to: ASSY_RAT from: 1 to: 412

1 ATGTCCAGCAAAAGGCTCGGTGCTGAGTACAGTGGCGCTGAGAC 50
|||||
1 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspTh 17
51 CTGCTGATCTCTCTGCTGCTGAGAGCAAGAGGCTATGACGTATGCT 100
|||||
17 rsercysileleuvaltyrleuylsgluuvalinglyltyrvalleialat 34
101 ATCTGGCCAAATGTCGCGCAAGAGACCTTCAGAGAGCCAGAGAGA 150
|||||
34 ylleuAlaasnllleglylnylsgluaspheglugluAlaargylsly 200
151 GCCTGAGAGCTTGGGGCAAAAGGTTTCATGATGATGATGATGAT 250
|||||
51 AlaLeuylsleuuglyAlaTylsValPheilegluaspvalserlysgl 67
201 GTTGTGGAGAGTTCATCTGCGCCCATCCAGCCAGCCAGCTGATG 250
|||||
67 uphevalglugluupheiletrproAlaValInserSerAlaLeuTyr 84
251 AGCAGCGTACTCTCGTGGCACTCTCTTCCAGGCGCTGATGCGCCG 300
|||||
84 lnsparlytyrleuendlythrserleuAlaargprocyllleAlarg 100
301 AAACAGTGGAAATGCGCCAGCGGAGGAGGCGCAAGATATGTCCTCC 117
|||||
101 LysgluValgluileAlaInlArggluInllylAlaTylValSerHisgl 117
351 CGCCAGCGAAAGGAGAGATGATGATGATGATGATGATGATGATG 134
|||||
117 yAlaThrcllylsglylAsnAspgluValArgPhegluLeuThrCysTyls 140
401 CACTGGCCCCCAATPAAAGTCATGCTCCCTGGAGAGATGCTGATTC 150
|||||
134 erleuAlaProgluInllylValleAlaProtrpArgMetArgGluPhe 500
451 TCAACCGGTTCAAGGCGCCATGATGATGATGATGATGATGATGAT 167
|||||
151 TyrAsnArgPheylsglylArgAsnAspLeuMetCgluTylAlaTylglnhl 167
501 CGGATTCCTCCATCCGCGTACCTCCCAAGAACCGGTGAGAGATGAGA 550
|||||
167 sgllyleProleProvalThrProlyserProtrpSerMetAspGluA 184
551 ACCATGACATCAGCTACGAGGCTGATGATGATGATGATGATGATG 200
|||||
184 snleuMetHisIleSerTylgluAlaaglylleuGlnAspProlyAsn 200
601 CAAGGCGCTCAGGCTCTACACAGAAAGCCAGAGCCAGAGCAAGCCC 650
|||||
201 GlnAlaArgProgluLeuTylThrlylGlnAspProAlaTylsAlaPr 217
651 CAACACCCCTGACATTCGAGATCGAGTCAAAAAGGGTCCCTGGA 700
|||||
217 oAsnThrProAspValleuGluInllegluPheTylsGlyValProval 234
701 AGTGAACAAGCTCAAGATGGACACACACAGACCTCTTGAAGCTC 750

```

|||||
234 yvalThrasValLysaspGlyThrHisSerThrSerLeuAspLeu 250
751 TTCATGACCTGACAGAGTCCGGGCAAGCATGCGTGGCCGTATGTA 800
251 PheMetTyrLeuAsnGluValAlaGlyHisGlyValAlaYarGlyLea 267
801 CATCGGAGAAACCGCTTCATTGGAATGATGATCCGAGTATGACAGA 850
267 pIleValGluAsnArpHeIleGlyMetLysSerArGlyIleYrGly 284
851 CCCGAGCAGCACCATTACCTTACATGCTATTAACATCGAGCCCTC 900
284 htrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAla 300
901 ACCATGACCGGCAAGTCCGCAAAATCAACAGCCCTGGCTGAATT 950
301 ThrMetAspArgGluValAlaArgLysIleLysGlnGlyLeuLys 317
951 TCGTGAAGCTGGTATACCGGTTTACGGCCCTAGCCCTGAGTGAATT 1000
317 eAlaGluLeuValTyrThrGlyPheThrPheSerProGluCysGluPhe 334
1001 TCCGCACTGACATCCGCAAGTCCGAGAGGAGTGAAGGAAAGTGCAG 1050
334 aAlaArgHisCysIleLysPheSerGlnGluValGluGlyLysValGln 350
1051 GTTCCTCTCAAGGCGCAGGTGTACATCTCGCCGCGGAGTCCCACT 1100
351 ValSerValPheLysGlyGlnValTyrIleLeuGlyArgGluSerPro 367
1101 GTCTCTACATGAGAGCTGTGAGCATGAGCTGAGGATGATG 1150
367 uSerLeuTyrAsnGluGluLeuValSerMetAsnValGlnGlySer 384
1151 AGCCACTATGACACCGGTTTCATCAATCAATCCCGAGCTGAG 1200
384 luoProIleAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeu 400
1201 GATATCATGCTCTCCAGAGCAAGTCAATGCGCAA 1236
401 GluTyrHisArgLeuGlnSerLysValThrAlaLys 412

seq_name: swissprot_39: ASSY_MOUSE

seq_documentation_block:
ID ASSY_MOUSE STRAND: PRT: 412 AA.
AC P16460;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUL-1999 (Rel. 15, Last sequence update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULINE--ASPARTATE
DE LIGASE).
OS ASS OR ASS1.
SN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_taxid=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=DKA/2J; TISSUE=Liver;
RA MEDLINE=91216457; PubMed=1708740;
RT Surr L.C., Beaudet A.L., O'Brien W.E.;
RT "Molecular characterization of the murine argininosuccinate
RT synthetase locus.";
RL Gene 99:181-189(1991).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULINE + L-ASPARTATE = AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M31690; AAA37266.1; -
DR EMBL; M31692; AAB60707.1; -
DR EMBL; M31694; AAB60708.1; -
DR EMBL; M31693; AAB60708.1; JOINED.
DR EMBL; M31695; AAB60708.1; JOINED.
DR PIR; M31702; AAB60706.1; -
DR MGD; MGI:88090; Ams1.
DR InterPro; IPR001518; Arginosuc_synth.
DR Pfam; PF00764; Arginosuc_synth.1.
DR ProDom; PD003544; Arginosuc_synth.1.
DR PROSITE; PS00564; ARGININOSUCCIN_SYN_1; 1.
DR PROSITE; PS00565; ARGININOSUCCIN_SYN_2; 1.
KW Arginine biosynthesis; Urea cycle; Ligase; ATP-binding.
FT NP_BIND 115 123 ATP (POTENTIAL).
SO SEQUENCE 412 AA; 46584 MW; A8F3AFDDFBAEF6A CRC64;

alignment_scores:
Quality: 2073.00 Length: 412
Percent Similarity: 98.786 Gaps: 0
Percent Identity: 96.117

alignment_block:
US-09-775-693-1 x ASSY_MOUSE
Align seg 1/1 to: ASSY_MOUSE from: 1 to: 412

1 ATGTCCAGCAAGAGCTCCGTGTTGCGCTACAGTGGCGGCTGAGAC 50
1 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspThr 17
51 CTGCGTCACTCCGTCGTGGCTGAGAACACAGCTATGACCTCATGCTT 100
17 rSerCysIleLeuValAlaThrPheLysGlnGlyValTyrAspValIleLeu 34
101 ATCTGGCCCAACATTTGCCAGAGAGACTTCGAGGAAGCCAGAGAAAG 150
34 YrLeuAlaAsnIleGlyGlnLysGluAspPheGlnGluAlaArgLys 50
151 GCACGTGAAGCTTGGGCGCAAAAGGTCTTATTGAGAGTGCAGAGGA 200
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerLysG 67
201 GTTGTGAGAGAGTTCATGCTGGCGCCGATCCAGTCCAGCCGACGTATG 250
67 uPheValGluGlnuPheIleThrProAlaValGlnSerSerAlaLeuTyrG 84
251 AGACCGCTACCTCTGGGACCTCTCTGACGAGCCCTGATGCCGCCG 300
84 luoAspArgTyrLeuLeuGlyThrSerLeuAlaAlaProCysIleAlaArg 100
301 AAACAAGTGAATCCGCCAGCGAGGAGGCGCAAGTATGTGCCACAG 350
101 ArgGlnValGluIleAlaGlnArgGlnGlyAlaLysTyrValSerHis 117
351 CGCACAGGAAGGGAGACGATCAGGTCCGTTTGGACCTGATGCTTACT 400
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuThrCysTyr 134
401 CACTGCCCCCAGATAAAGTCAATGCTCCCGAGAGAGTGCCTGAATT 450
134 eLeuAlaAlaProGlnIleLysValIleAlaProIlePheMetProGluPhe 150
451 TACAACCGTTCAAGGCGCAGATGACTGATGAGTACGCAAGCAACA 500

```



```

|||||
151 TyAsnArgPheLysGlyValArgAsnSpmLeuMetGlyTyrAlaLysGlnH 167
501 CGGGATTCCCATCCCGTCACTCCAGAACCCGCGGAGCATGTGATGAGA 550
167 sGlyLeuProLleProValThrProLysSerProTrpSerMetLysGln 184
551 ACCTGATGACATCAGTACAGAGGCGGGAATCCTGAGAACCCCAAGAC 600
184 snLeuMetHisLleSerTyrGlnAlaGlyLleLeuGlnAsnProLysAsn 200
601 CAACGGCTCCAGGCTCTCTACAGAAACCCAGAGCCCAAGCCCAAGCCC 650
201 GlnAlaProProGlyLeuTyrTyrTyrTyrGlnAsnProAlaLysAlaPr 217
651 CAACACCCCTGACATCTCTGAGATCGAGTTCAAAAAAGGGCTCCGTGCA 700
217 GAsnSerProAspValLeuGlnLleGlnPheLysGlyValProValL 234
701 AGGTGACCAACGTCAAGATGACGACACCCACCAAGCCTCTTGAGAGTCT 750
234 yValThrAsnLleLysAspGlyThrThrArgTyrThrSerLeuGlnLeu 250
751 TTCATGTACCTGAAGATCGCGGCAAGCATGCGCGGTGCGCTATGGA 800
251 PheMetTyrLeuAsnGlnValAlaLleLysHisGlyValGlyArgLys 267
801 CATGTGAGAAACCGCTTCAATTGAAATGAATGACCCGAGGTATCTACAGA 850
267 PLeuValGlnAsnArgPheLleLysMetLysSerArgLysTyrGlnT 284
851 CCCAGAGAGGACACATCCTTACATGCTCATTTAGACATCGAGGCTTC 900
284 hrProAlaGlyThrLleLeuTyrHisAlaHisLeuAsnPrLleGlnAlaPhe 300
901 ACCATGACCGGGAAGTGGCCAAATCAAAACAGAGCCTGCGCTTGAAAT 950
301 ThrMetAspArgGlnValArgLysLleLysGlnGlyLeuLysPsh 317
951 TGGTGGCTGGGTATACCGGTTTACCGGCTTACGCGCTGAGTGTGAATTG 1000
317 eAlaGlnLeuValLysThrGlyPheThrPheTrpHisSerProGlnLysGlnPhe 334
1001 TCCGCCACTGATCCGCAAGTCCCAAGGAGATGAGTGAAGGAAAGTGCAG 1050
334 aAlaArgHisCysLleGlnLysSerGlnGlnArgValGlnGlyLysValGln 350
1051 GTGTCCGTCTCAAGGCGGAGGTGATACCTCTGCGCGGAGTCCCACT 1100
351 ValSerValPheLysGlyLysValLysLleLeuGlnArgGlnSerProLe 367
1101 GTCTCTCTACATGAGAGAGCTGTGAGCATGAAAGTGCAGGTCATTATG 1150
367 uSerLeuTyrAsnGlnLysValSerMetLysValGlnGlyAspTyrG 384
1151 AGCCAATGATGCCACCGGCTTCAATCAATCAATCTCCCTGAGGCTGAG 1200
384 lUpProLleAspAlaThrGlyPheLleAsnLleAsnSerLeuArgLeuLys 400
1201 GAATATCATGCTCTCCAGAGAGTCACTGCAAA 1236
401 GlnTyrHisArgLeuGlnSerLysValThrAlaLys 412
seq_name: SwissProt_39: ASSY_BOVIN
seq_documentation_block:
ID ASSY_BOVIN STANDARD; PRT; 412 AA.
AC P14568;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
LIGASE).

```

```

GN ASS.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90046714; PubMed=2813370;
RA Dennis J.A., Healy P.J., Beaudet A.L., O'Brien W.E.;
RT "Molecular definition of bovine argininosuccinate synthetase
RT deficiency."
RT Proc. Natl. Acad. Sci. U.S.A. 86:7947-7951(1989).
RL -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC PATHWAY.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- DISEASE: DEFECTS IN ASS ARE THE CAUSE OF A BOVINE FORM OF
CC CITRULLINEMIA.
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M26198; AAA30388.1;
CC PIR: A33986; AUROS.
CC InterPro: IPR001518; Arginosuc_synth.
CC Pfam: PF00764; Arginosuc_synth; 1.
CC ProDom: PD003544; Arginosuc_synth; 1.
CC PROSITE: PS00564; ARGININOSUCCIN SYN.1; 1.
CC PROSITE: PS00565; ARGININOSUCCIN SYN.2; 1.
CC DR Arginine biosynthesis; urea cycle; ligase; ATP-binding.
CC NP_BIND 115 123
CC FT SEQUENCE 412 AA; 46417 MW; 6F7AC7F445EED086 CRC64;
SO

```

```

alignment_scores:
Quality: 2069.00 Length: 412
Ratio: 5.059 Gaps: 0
Percent Similarity: 99.272 Percent Identity: 95.388

```

alignment_block:

us-09-775-693-1 x ASSY_BOVIN ..

Align seg 1/1 to: ASSY_BOVIN from: 1 to: 412

```

1 ATGTCCAGCAAAAGGCTCCGGTTCGCTTACAGTGGCGGCTTGACAC 50
|||||
1 MetSerGlyLysGlySerValValLeuAlaLysSerGlyGlyLeuAspH 17
51 CTCGTGATCTCCGCTGTGGTGAAGAAACAAGCTATGAGCTATGCGCT 100
17 rSerGlyLleLeuValTrpLeuLysGlnGlnGlyTyrAspValLleAla 34
101 ATGTGGCCACATFTGGCCAGAAAGAACTTGAAGAGCCAGAGCAGAGAC 150
|||||
34 yLeuAlaAsnLleGlyGlnLysGlnAspPheGlnGlnAlaLysLys 50
151 GCACGTGAAGCTTGGGCGCAAAAGGTGTTTCATTTAGAGATGTCACAGGA 200
|||||
51 AlaLeuLysLeuGlyAlaLysLysValPheLleGlnAspLleSerLysG 67
201 GTTTGTGAGAGACTTCATCTGGCGGCGCATCCAGTCCAGCGCATGTATG 250
|||||
67 uPheValGlnGlnPheLleTrpProAlaLleGlnSerSerAlaLeuTyrG 84
251 AGAACCCGTACTCTCTTGAGGACCTCTTTCGCAAGGCCCTGCATGCCCGC 300

```

384 luprVlaAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeulys 400
1201 GAAATCATCGCTGCACAGCAAGTCACTGCCAA 1236
|||||
401 GIuTYrHISArgLeuGlnAsnIleValAlaThrAlalys 412

seq_name: SwissProt 30.1000

seq_documentation_block:

AC 097069; Q9VT41: 419 AA.

DT	30-MAY-2000	(REL. 39, Created)

20 AUG-2001 (Rel. 40, Last annotation update)

ASPARTATE LIGASE). (EC 6.3.

Drosophila melanogaster (Fruit fly)

Pterynota; Neostomatopoda; Arthropoda; Tracheata; Hymenoptera; Meladzoa; Noctuidae

Phyllophaga; Drosophilidae; Drosophila

三

STRAIN=BERKELEY.

Preitter B., Knafels J., Marti
Palazzo M. T.

complete sequence of the *Antennapedia* complex

[2] <http://www.fishbase.org/Species/SpeciesBank/DBSJ>

C STRAIN=BERKELEY:

Adams M D. *Colobiont*. PubMed=10731132;

George D. Scherer S.E., Lt. P.W. Hosk

Sutton G.G., Wortman J.R., Yandell M.D., Asndourne

Wan K.H., Doyle C., Baxter F.C., Blazej R.G., Che

Ballew R.M., An H.-J., Andrews-Pf...

Borokova D. A., Benos P. V., Berman B. P., Bhandari

Busam D.A., Butler H

de Pablos B., Delcher A., Davenport L.

Durbin K. J., Downes M., Dugan-Rocha S.

Gard N.S., Gelbart

Harris N.L., Harvey D., Heiman T., Guan

Jalali M., Kalush F., Karner G., Nowland J., Wei M.-H.

Lasko P., Kraft C., Kravitz S.

Mattei B., McIntosh T.C., McLeod M P

Mount S.M., Moy M., Murphy B., Morris

Palazzolo M., Pittman G., Nusskern D.

Shue B.C., Wilmington K., Saunders R.D.C., Sche

Strong
Stapleton M.
Spradling A.C.,
Sprei E.,

wang Z.-Y., Wassarman D.A., Weinstock G.M., Veller E., W

Ye J., Yeh R.-F., Zaveri T. S., Worley K. C., Wu D.,

Gibbs R. A., Zhong E. N., Zhong W., Zhou X., Zhu S.

the genome sequence of *Drosophila melanogaster*

-!- CATALYTIC ACTIVITY: ATP + L-CYSTEINE

... D-ARGININOSUCCINATE.

```

CC CC      -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC CC      PATHWAY.
CC CC      -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC CC      -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation. Its
CC CC      European Bioinformatics Institute. There are no restrictions on its way
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC CC      or send an email to license@isb-sib.ch.
CC CC      -----
CC CC      EMBL: AE001574; AAD19816.1; -
CC CC      DR EMBL: AE003674; AAF54103.1; ALT-SFO.
CC CC      DR EMBL: FBGN0026565; BG:DS00004.14.
CC CC      DR FlyBase: IPR001518; Arginosuc_synth.
CC CC      DR InterPro: IPR001518; Arginosuc_synth; 1.
CC CC      DR Pfam: PF00764; Arginosuc_synth; 1.
CC CC      DR ProDom: PD003544; Arginosuc_synth; 1.
CC CC      DR PROSITE: PS00564; ARGININOSUCCIN_SYN_2; 1.
CC CC      DR PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1.
CC CC      DR Arginine biosynthesis; Urea cycle; Ligase; ATP-binding.
CC CC      KW Arginine biosynthesis; Urea cycle; ATP (POTENTIAL).
CC CC      NP BIND 114 122
CC CC      FT sequence 419 AA: 46594 MW: 0566A10A329BE513 CRC64:
CC CC      SEQUENCE

```

```
alignment_block: ..  
US-09-775-693-1 x ASSY_DROME  
  
Align seg 1/1 to: ASSY_DROME from: 1 to: 419  
  
.....CGTGGGCTGTGGACACTCTGTGCAT 59
```

[illegible]

```

153 PheingnlgyArgglnAspLeuIlealIyAlaIglnglnIhIsglyIleI 169
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
154 CATCCCGGTCACTCCCAAGAACCCGTGGACCATGATGAGAACCATCATGC 559
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 valIserIAlaIyProIleAlrPrOTrPserThIAspAlaAsnIleLeuN 186
169 uAlserIAlaIyProIleAlrPrOTrPserThIAspAlaAsnIleLeuN 609
560 ACATCAGCTACGAGCGCTGGAAATCCGGAGAACCCCAAGAACCAAGCGCC 202
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
186 IsIleSerIyIgluSerIyIleLeuGlnAspProAsnThIValIAlPro 219
186 IsIleSerIyIgluSerIyIleLeuGlnAspProAsnThIValIAlPro 656
610 CCAGAGCTCTACACAGAACCCAGAACCA...GCCAAGCCCCAACAC 656
||||:||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 GluAsnLeuTyIgluMetIhValAlaPrIoleuNlrArgAlaIProIArgS 219
203 GluAsnLeuTyIgluMetIhValAlaPrIoleuNlrArgAlaIProIArgS 706
657 CCCGACATCTCTGAGATCGAGTTCACAAAAAAGGGTCCCTGTGAGAGTGA 706
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 PrIovalhIsIleuValIleIglInPheAspArgIleuProSerSerValG 236
219 PrIovalhIsIleuValIleIglInPheAspArgIleuProSerSerValG 756
707 CCACGTCACAGATGGAGCCACCACCCACCAAGCACTCTTGGAGCTTCATG 756
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 lAspIleuPrOglIyIAlrValIyTrIhIySPrIoleuNlrIgluMetIleuAsp 252
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
757 TACCTTAACAGAGATGCGCGGCGACAGATGCGTGGCGCCGTATTCATTCG 806
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
253 PheLeuAsnIySleuIyIleSerIyIleIyAlaIyIleAspIleAlIle 269
807 GGAAGACCGCTCATTTGAGATAGAGTCCCGAGGTATCTACAGAACCCAG 856
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
269 lgluAsnArIghPheValIglIleuIySerArIglIyAlIyIgluIhPrOg 286
857 CAGGCACCATTCCTTACCATGCTCATTTAGACATCGAGGCGCTTCACCATG 906
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
286 lYgIyThIleuPheAlaIAlaIhIsIgluAspIleuGlnuAlPheAlaIleu 302
907 GACCGGGAAGTGCGCAAAATCAACACAGAGCGCTGGCTGAATTCGTGA 319
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
303 AspArIyGlnuAlIleuArIhIySgluValIleuArIghAspArIghIleAlaAs 319
957 GCGGGGTATTCACCGGTTTACGGCTCAGCGCTGAGTGTGAATTCGCCG 1006
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
319 pIyIyAlIyTyAsnIyIghPheTrIphSerPrOgluAlaIleTyIAlaIygl 336
1007 ACtGACtGCGCAGAGTCCAGAGAGCGAGTGGAAAGGAAAGTGCAGGCTGCC 1056
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
336 yScyIleIgluIleAlIgluIghIAlrIyAlSerIyIySAlIhIyValIglu 352
1057 GTCTCTCAAGGCGCCAGGTGTCATCTCTCGCGCGGAGTCCCCACTG... 1101
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 leuAlaPrOglIyIyCysArIghAlaIleAlaArIghIySAlaIAlaIyAspIyA 369
1102 ...TCTCTTACAATGAGAGAGCTGGAGCATGACGTGACGGGTGAT 1147
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
369 lgluAlaIleuTyAsnIyIghIleuValIleSerMetAspAlaIhIsIyglY 386
1148 ATGAGCGCACTGATGCTCCAGCGGTTTCATCAACATCAATTCCTCGAGCTG 1197
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
386 yIyAlPrOglInAspAlaIyglIyPheIleAlaIleAsnAlaIyAlrArg 402
1198 AAGGAATATCATGT 1212
||||:||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 ArgIghIhIyAlrArg 407
seq_name: SwissProt_39:ASSY_SCHPO
seq_documentation_block:
ID ASSY_SCHPO STANDARD; PRT; 410 AA.
AC 094354;
DC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ARGININOSUCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
LIGASE).

```

```

GN  SPAC428.05C.
OS  Schizosaccharomyces pombe (Fission yeast).
OC  Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC  Schizosaccharomycetales; Schizosaccharomycetaceae;
ON  NCBI_TaxID=4896;
RX  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=972;
RA  Lyne M., Rajandream M.A., Barrell B.G., Baker S., Mungall K.;
CC  Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC  -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC  PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC  -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC  PATHWAY.
CC  -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  use by non-profit institutions. There are no restrictions on its
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
DR  EMBL; AL034382; CAA22280.1;
DR  InterPro; IPR001518; Argininosucc_synth.
DR  Pfam; PF00764; Argininosucc_synth.
DR  ProDom; PD003544; Argininosucc_synth.1.
DR  PROSITE; PS00564; ARGININOSUCCIN_SYN_1.1.
DR  PROSITE; PS00565; ARGININOSUCCIN_SYN_2.1.
KW  Arginine biosynthesis; urea cycle; Ligase; ATP-binding.
FT  NE_BND 116 124
SQ  SEQUENCE 410 AA; 46084 MW; 1B3873A4DCDC21 CRC64;

alignment_scores:
Quality: 1083.00
Ratio: 3.384
Percent Similarity: 81.013
Length: 395
Gaps: 4
Percent Identity: 54.430
US-09-775-693-1 x ASSY_SCHPO
Align seg 1/1 to: ASSY_SCHPO from: 1 to: 410

22 GTTGTGCTTACAGTGGCGGCTGAGCACTTCGTCATCTCGTGGGCT 71
9 ValLeuAlaIyrSerIyLylLeuAspThrSerCysIleLeuAlaIyr 25
72 GAAGGAACAAGGCTATGAGCTGCTATTCGTCGTCGTCGTCGTCGTC 71
25 uirleuglulgllytrpgluvalillecystyrmctalaasnaValglgylng 42
122 AGGAGACTTCAGGAGAACCCAGGAAGAGGACTGAGCTTGGGGCCAAA 171
42 IugluAspTrpAspAlaAlaIargIuAlaIeuValIeuValglYAlaIys 58
172 AAGGTTCATTCAGATGATGACAGGAGGATTGTGAGAGGATTCACCTG 221
59 LysValIyrValIgluAspLeuAargIuIupheIleAsnaSphrYAlil 75
222 GCCGGCATTCAGTCCAGCGCACTATGAGAGACCGCTACCTCTCGGCA 271
75 eProAlaIaIagIuAlaIaIaIaIeIyrcIuAsnaValIyrLeuLeuGlY 92
272 CCTCTTCCAGGCCCTGTCATCGCCGCAACAAGTGAATAATCGCCAG 321
92 hsrIeuAlaIarproIleIleAlaIarIargIuIleIglIleAlaIu 108
322 CGGAGGGGGCCAAAGTATGTGTCCACGGGGCCAGGAAGGGAAGCA 371

```

seq_name: SwissProt_39:ASSY_YEASr
seq_documentation_block:

Wed Feb 13 07:36:07 2002

us-09-775-693-1.rsp

ID ASSY_YEAST STANDARD: PRT: 420 AA.
AC P22768;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE-ASPARTATE
LIGASE).
GN ARG1 OR YOL058W OR O1228.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91071613; PubMed=2123815;
RA van Vliet F., Crabbeel M., Boyen A., Tricot C., Stalon V., Falmagne P.,
Nakamura Y., Baumberg S., Glansdorff N.,
RT "Sequences of the genes encoding argininosuccinate synthetase in
Escherichia coli and Saccharomyces cerevisiae: comparison with
methanogenic archaeobacteria and mammals.";
RL Gene 95:99-104(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / FY73;
MEDLINE=96381248; PubMed=8789261;
RA Mannhaupt G., Vetter I., Schwarzlouse C., Mitzel S., Feldmann H.;
RT "Analysis of a 26 kb region on the left arm of yeast chromosome XV.";
RL Yeast 12:67-76(1996).
RN [3]
RP SEQUENCE OF 1-57 FROM N.A.
RX MEDLINE=8823481; PubMed=2897249;
RA Crabbeel M., Seneca S., Devos K., Glansdorff N.;
RT "Arginine repression of the Saccharomyces cerevisiae ARG1 gene.
Comparison of the ARG1 and ARG3 control regions.";
RL Curr. Genet. 13:113-124(1988).
CC -1- FUNCTION: IN YEAST, AS CAN HAVE A CATABOLIC FUNCTION SINCE IT
ALLOWS EFFICIENT UTILIZATION OF CITRULLINE VIA ARGININE AND THE
REACTIONS INVOLVED IN THE ARGINASE PATHWAY.
CC -1- CATABOLIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
PATHWAY.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M5237; AAA34437.1;
DR EMBL: X91067; CA662528.1;
DR EMBL: Z74800; CA99067.1;
DR EMBL: X07070; CA30106.1;
DR PIR: J00779; AUBRS.
DR SGD: S0005419; ARG1.
DR InterPro: IPR001518; Argininosucc. synth.
DR Pfam: PF00764; Argininosucc. synth. 1.
DR ProDom: PD003544; Argininosucc. SYN_1.
DR PROSITE: PS00564; ARGININOSUCCIN_SYN_2; 1.
DR PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1.
KW Arginine biosynthesis; Urea cycle; Ligase; ATP-binding.
NP_BIND 114 122
FT CONFLICT 27 28
FT CONFLICT 48 49
FT CONFLICT 61 64
FT CONFLICT 169 169
FT CONFLICT 316 316
SQ SEQUENCE 420 AA: 46939 MW: 785309DD58443860 CRC64;

alignment_scores: Length: 414
Quality: 1069.00 Gaps: 4
Ratio: 3.351

Percent Similarity: 77.053 Percent Identity: 50.483

alignment_block:
US-09-775-693-1 x ASSY_YEAST

Align seg 1/1 to: ASSY_YEAST from: 1 to: 420

7 AGCAAGGCTCCGCTGTTCTGCTACAGTGGCGGCTGAGACCTGCTG 56
18 LLELEALATRPLEULEAUSPDLNGLYTGTGVALVALALAPHEVELA 18
2 SERLYSGLYVALCYSLLEUALATYISERLYGLEYLEUSPHTSERVAL 106
57 CATCTCTGCTGTGCTGAAGAACGCTATGACGTCATTCCTGCTGCTG 35
18 LLELEALATRPLEULEAUSPDLNGLYTGTGVALVALALAPHEVELA 156
107 CCACATTTGGCCAGAGAACGCTTGGAGAACCCAGAGAACGAGCAGCTG 51
35 LAASNVALGGLGGLNGGLUASPHEASPALALALGSLUYSALALEU 206
157 AAGCTTGGGCGCAAAAGGTTCATTGAGATGTCAGAGGAGATTGCT 68
52 LYSILEGLYALACYSLYSPHEVALCYVALASPCYSARGGLASPEVEA 256
207 GAGAGATTCATCTGGCGGCTCCAGCTCCAGCTCCAGCTGATGAGACC 85
68 LLYASPILEULEUPHEPROVALVALGVALASVALVALVALYGLUASPV 306
257 GCTACCTCTGGGACGCTCTCTGCGAGGCTCCGCTGCTGCTGCTGCT 101
85 ALLYRLEULEUGLYTRSERLEUALARGPROVALLEALALYSALAGLN 356
307 GTGCAATTCGCCAGCGGAGGGGCGCAAGTATGTGTCACGCGCCAC 118
102 LLESPALALALYSGLNGGLYSPHEALVALSERHLSGLYCYSTH 406
357 AGGAAGGAGGAGCATGATGCTCGTTGAGCTCGCTGATGATGATGAT 135
118 TELYSLYSLYASNPDLNGLATRPHEGLYLEUSERPEHYRALLALEU 456
407 CCCCCAGATTAAGTGCATGCTCTCTGCGAGATGCTGATTTACAC 151
135 YSPROASPVALLYSCTSLTERHPROTIRPMPETPROGLUPHEPGLU 506
457 CGGTTCAAGGCGCCGATGATGATGATGATGATGATGATGATGATGAT 168
152 ARGPHALAGLYARGLYASPLEULEUSPYRALALAGLNLYSGLYTL 556
507 TCCCATCCCGCTCACCCCAAGAACCCGTCGAGCATGATGATGATGAT 185
168 EPROVALALAGLNLHYSLALALYSPTOTRPSERTHASPDLUSNGLA 606
557 TGCACATCAGCTACGAGGCTGGAATCCGAGAACCCCAAGAACCAAGC 201
185 LAHLSLESETYGLUALAGLYLLEUGLNASPROASPTHTHPTHTHPT 656
607 CCTCAGGCTCTCTACAGAACGAGACCCAGACCCAGAACCCCAAGAC 218
202 PROLYASPMETTRPYSLEULLEVALASPPROMETASPALAPROASPL 706
657 CCCTGACATTTCTCGAGATCGAGTTGCAAAAAGGGTCCCTGTAAGGTA 235
218 NPROGLNASPLEUTHTHLEASPHREGILUATAGLYLEUPROVALYSL 744
707 CCAAGCTCAAGATGACACCAACCAACCAACCAACCAACCAACCAAC 251
235 HTYRTHRSPASNLSTHSTHSTHSTHSTHSTHSTHSTHSTHSTHSTH 794
745 GAGCTCTCATCTGACTGAGAACGATCCGCGGCAAGCATGCGTGGCGG

```

252 AspvAlpheleuAlaAlaSerAsnleuAlaArgAlaAsnGlyValGlyArg 268
795 TATTGACATGTCGAGACCGCTTCATTGGAATGAATGCCGAGAGTTC 844
268 GIIleAspIleValGIIAspArgTyrIleAsnleuLysSerArgGlyGly 285
845 ACAGAGACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 894
285 YrIleuIleAlaProleuThrValleuArgLysAlaHisValAspLeuGlu 301
895 GCGTTCACCATGACCGGAGAGTGCAGCAAAATCAACACAGGC...CTGGG 941
302 GlyLeuThrLeuAspLysGlyValAlaGlyIleAlaArgAspSerPheValI 318
942 CTGGAATTTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 991
318 ProAsnThrSerArgLeuIleuIleuIleuIleuIleuIleuIleuIleu 1041
992 GTGAATTTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1041
335 YsgIuTyrIleArgSerMetIleGlnProSerGlnAsnSerValAsnGly 351
1042 AAAGCTGAGGTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1091
352 ThrValArgValArgLeuTyrGlyValAsnValIleIleLeuGlyArgSe 368
1092 GTCCCACTG...TCTCTCAACAATGAGAGAGCTGTGATCATCAAGCTGC 1138
368 rThIuTyrThrGluTyrAspProThrGluSerSerMetAspGluL 385
1139 AGCGTCAATTTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1188
385 euThrGlyPheLeuProThrAspThrIleAlaIleAlaIleGlnAla 401
1189 CTCAGCCTGAGCAATATCATCTCTCCAGACAGCAGCAGTCACT 1230
402 IleaGlyIleuLysTyrGly.....GluSerLysTyrThr 413

seq_name: SwissProt_39:ASSY_ARATH
seq_documentation_block:
ID ASSY_ARATH STANDARD PRT 523 AA.
AC Q9S2X3;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE ARGININOSUCCINATE SYNTHASE, CHLOROPLAST PRECURSOR (EC 6.3.4.5)
GN ARG24830 OR F617.40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Mambo R., Murphy G., Voickaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Meche R., Mueller M.,
RA Kreis M., Delseny M., Fuldomech P., Watson M., Schmidheini T.,
RA Reihert B., Portelle D., Perez-Alonso M., Bouty M., Bancroft I.,
RA Vos P., Heideisel J., Zimmermann W., Medler H., Ridley P.,
RA Langham S.-A., McCullagh B., Billam L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Baetens I., Kert R., Defoor E.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Braun M.,
RA Woolman P., Klein lankhorst R., Rose M., Hauf J., Koetter W.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gijlen J., Villarroel R., De Clercq R.,

```

```

RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Petre A., Rajandream M.-A., Lyne M., Benes V., Reclman S.,
RA Borkova D., Bloecher H., Scharfe M., Grimm M., Loehert T.-H.,
RA Dose S., de Haan M., Maatse A., Schaefer M., Mueller-Auer S.,
RA Neumann S., Argilou A., Vitale D., Liguori R., Piravadi E.,
RA Massenot O., Quigley F., Clabaud G., Wundelstein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lechay A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casabonera S.,
RA Gibbons F., Weber N., Vandenbol M., Bagues M., Teyl J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedbia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sehkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud B., Miller A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vili D., Shekher M., Matero A., Shah R.,
RA Granat S., Shoddy N., Hasegawa A., Rodriguez M., Hoffman J., Till S.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.,
RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RA thaliana."
RA Nature 402:769-777(1999).
RP CONCEPTUAL TRANSLATION.
RL Lemke C.;
RU Unpublished observations (JUN-2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: PENULTIMATE STEP IN ARGININE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO WRONG EXON
CC PREDICTIONS FROM THE GENOMIC SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: AL049657; CAB41123.1; ALT_SEQ.
DR EMBL: AL161562; CAB79393.1; ALT_SEQ.
DR InterPro: IPR001518; Arginosuc_synth.
DR Pfam: PF00764; Arginosuc_synth.
DR Prodom: PD003544; Arginosuc_synth.
DR PROSITE: PS00564; ARGININOSUCCIN SYN. 1; 1.
DR PROSITE: PS00565; ARGININOSUCCIN SYN. 2; 1.
KW Arginine biosynthesis; Ligase; ATP-binding; Transit peptide;
KW Chloroplast.
FT TRANSIT 1 ?
FT CHAIN 1 ?
SO SEQUENCE 523 AA; 57249 MW; 97B82A5A718969C8 CRC64;

```

alignment_scores:

```

Quality: 911.50
Ratio: 3.018
Percent Similarity: 76.263
Length: 396
Gaps: 4
Percent Identity: 45.707

```

alignment_block:

```

US-09-775-693-1 x ASSY_ARATH

```

```

402 hrlleuphealaalavalglnlulengluserleuthrleuasparg 418
913 GAAGGGCGCAAAATCAACACAGCGCTGGCGTGAATTGCTGACGTCT 962
||| :::::::::::::: :::::::::::::: ::::::::::::::
419 Gluserlleglnvallysasprrheulauleuysrlyalaglmetva 435
963 GTATACCGGTTTACGCGCTAGCCCTGACTGTGATTTGTCGCCACTGCA 1012
||||| :::::::::: :::::::::: ::::::::::
435 ltryalaglalygrttrpheapsproleuarglusermetaspalaphe 1062
1013 TCCGCAAGTCCAGGACGAGGTGAAGGAAAGTGCAGCTGCTCCGCC 468
::| | | | | ::| | | | | ::| | | | | ::| | | | |
452 etgluyslserlthrgrluthrthrlyservalthrleuylsleuty 1112
1063 AAGGCGAGGTGTACATCCTCGCCGCCGAGTCCCTCGTCTCTACAA 1162
||||| ::| | | | | ::| | | | | ::| | | | | ::| | | | |
469 lylslyservalservalthrlyarglinserserproanserleuty 1162
1113 TGAGAGCTGTGTACATGACATGACGTGACGAGGTGATTTGAGCCATG 1200
485 gglinspileserserphedgluglyserglulserglulserglnal 502
1163 CCACGCGGTTCATCAACATTCATCCCTCAGGCTGAG 1200
502 lahalaglypheleargleuertyclyleuprometlys 514
seq_name: SwissProt_39:ASSY_BACSU
ID ASSY_BACSU STANDARD: PRT: 403 AA.
seq_documentation_block:
AC 034347;
AC 15-JUL-1999 (rel. 38, Created)
DT 15-JUL-1999 (rel. 38, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE-ASPARTATE
DE LIGASE).
GN ARG-
OS Bacillus subtilis.
ON Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98048467; PubMed-9387221;
RA Lapius A., Galleron N., Sorokin A., Enrich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
in the 200 kb rmb-dnaB region.";
RL Microbiology 143:3431-3441(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
L-PROLYSOPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: PENULTIMATE STEP IN ARGININE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOPENTAMER (BY SIMILARTY).
CC -1- SIMILARTY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Use by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF0088220; AAC00320.1; -
DR EMBL: Z69119; CAB14923.1; -
DR EMBL: Z69118; CAB14905.1; -
DR Subtilist: BG12570; argG.
DR InterPro: IPR001518; Argininosuc_synth.
DR Pfam: PF00764; Argininosuc_synth.1.
DR ProDom: PD003544; Argininosuccin_Syn_1; 1.
DR PROSITE: PS00564; ARGININOSUCCIN_SYN_2; 1.
DR PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1.
KW Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.
KW Argininosuccinylase; BFE552A77386111F CRC64;
SQ

```

alignment_scores:

Quality: 818.00 Length: 400
 Ratio: 2.773 Gaps: 3
 Percent Similarity: 73.750 Percent Identity: 42.750

alignment_block:
 US-09-775-693-1 x ASSY_BACSU ..

Align seg 1/1 to: ASSY_BACSU from: 1 to: 403

```

1 ATGTCCAGCAAGAGCTCCGTTGCTTGGCTTACGATGAGGCGGCTGAGAC 50
1 MetSerGlnGlnLysValValLeuAlaTyrSerGlyGlyLeuAspThr 17
51 CTCGTGATCTCTGCTGCTGAGCAAGAGCTATGACGTATGCTATGCTT 100
17 rSerValAlaLeuTyrPheGlnGlnGlnGlyTyrAsnValIleAla 34
101 ATCTGCCAACATTTGGCCAGAGAGCTTCCAGAGCCAGCAAGAGAG 150
34 yScysLeuAspValGlyGlnGlyLysAspLeuAlaPheValGlnGln 50
151 GCATGAGAGCTTGGGCGCAAAAGAGCTTCTATGAGATGATGACAGGA 200
51 AlaLeuGlnValGlyAlaThrAsnSerTyrValIleAspAlaLysGln 67
201 GTTTGTGAGAGCTTCACTTGGCGGCGCATCCAGCCAGCCACTGATG 250
67 uPheAlaGlnAspTyrAlaLeuIleSerLeuGlnAlaIleThrMet 84
251 AGGACCGCTACTCTCGGCGCATCTCTTCCAGGCGCTCATGCGCCG 300
84 LuGlyLysTyrProLeuValSerAlaLeuSerAlaProLeuIleAla 100
301 AAACAAGTGAATGCGCCAGCGAGGCGCAAGTATGATGTCGCCAGG 350
101 LysLeuValGlnIleAlaGlnLysGlnAspAlaGlnAlaIleAla 117
351 CGCCAGCAAGAGAGAGAGATGATGATGATGATGATGATGATGATG 117
117 yGlyThrLysGlnLysAsnAspGlnValAlaPheGlnValSerLeu 134
401 CACGTGCGCCCGAGATTAAGATGATGATGATGATGATGATGATGAT 450
134 eTLeuAsnProAspLeuGlnValIleAlaProValAlaGlnLysP 450
451 TACAACCGGTTCAAGGCGCATGCTGATGATGATGATGATGATGAT 150
151 .....SerArgLysGlnGlnIleGlnLysAlaAlaSer 162
501 CGGATTCATCCATCCGCTCCTCCCAAGAGCCGAGGATGATGATG 162
162 gGlyIleProIleProIleAsnLeuAspSerProTyrSerIleAsp 179
551 ACCCTATGCATCAGCTCAGGAGGAGATGATGATGATGATGATGAT 600
179 sNLeuTyrPglYArgAlaAsnGlnGlyLeuGlnAspProPala 195
601 CAAGCGCTCCAGGCTCTACAGAGAGAGCCAGGAGCCAGGAGAGCC 650
196 AlaProGlnGlnLysAlaTyrAspLeuThrAlaProLeuLysThr 212
651 CAACACCGCTGATCTGATGATGATGATGATGATGATGATGATGAT 700
212 oAspThrProGlnValIleGlnIleAlaPheGlnGlnGlnValAla 229
701 AGGTGACCAAGCTCAAGAGAGAGCCAGGAGCCAGGAGCCAGGAG 750
229 eTLeu.....AspGlyValSerTyrSerLeuSer.....Gln 240
751 TTCATGTACTGAGAGAGAGTGGGCAAGCATGCGCTGAGGCGCTAT 800

```

```

241 IleLeuLysLeuAsnGlnMetAlaGlyAlaIleSerGlyAlaGlyLeu 257
801 CACGTGAGAGAGCCGCTTCTATGAGATGATGATGATGATGATGATG 850
257 pHisValGlnAsnArgLeuValGlyIleLysSerArgValAlaTyr 851
851 CCCAGCAGGAGCCATCTTACCATGCTTATGATGATGATGATGATG 900
274 yPheGlnAlaMetThrLeuIleValAlaIleLysGlnGlnAspLeu 901
901 ACCATGAGCCGAGAGAGTCCCAAAATCAACAAGAGCCAGGCTGAA 950
291 ThrLeuValLysGlnValAlaIleAspPheLysProIleIleGln 951
951 TGTGACAGCTGATGATGATGATGATGATGATGATGATGATGATG 1000
307 tSerGlnIleIleTyrAsnGlyLeuTyrPheSerProLeuLysAsp 1001
1001 TCCGCGCAGTGCATCCGCAAGTCCAGAGAGGAGAGAGAGAGAGAG 1050
324 euHisAlaPheLeuLysGlnLysGlnLysValIleThrLysIleVal 1051
1051 GGTCTGCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1100
341 ValLysLeuPheLysGlnHisAlaIleValGlnGlyArgLysSer 1101
1101 GTCTCTACATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1150
357 rSerLeuTyrAspGlnLysLeuAlaThrTyrThrLysAspAlaPhe 1151
1151 AGCAACTGATGCCACCGGTTTCATCAATCATCATTCCTCAGCGTGA 1200
374 sPHisAlaIleAlaIleGlyPheIleGlnLeuTyrProLysProThr 1200
seq_name: SwissProt_39:ASSY_BUCAI
seq_documentation_block:
ID ASSY_BUCAI STANDARD: PRT: 403 AA.
AC P57138;
DT 20-AUG-2001 (Rel. 40. Created)
DT 20-AUG-2001 (Rel. 40. Last sequence update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE-ASPARATE
DE LIGASE).
GN ARG-OR BU050
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RT Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT Buchnera sp. APS.
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARATE = AMP +
CC -1- PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions. There are no restrictions on its
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC EMBL: AP001118; BAB12773.1.
DR

```


DR InterPro: IPR001518; Arginosuc_synth.
 DR Pfam: PF00764; Arginosuc_synth; 1.
 DR ProDom: PD003544; Arginosuc_synth; 1.
 DR PROSITE: PS00564; ARGININOSUCCLIN_SYN_1; 1.
 DR PROSITE: PS00565; ARGININOSUCCLIN_SYN_2; 1.
 DR Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.
 KW SEQUENCE 403 AA; 45072 MW; F23B7024C8BDB147 CRC64;

alignment_scores:
 Quality: 807.50 Length: 396
 Ratio: 2.710 Gaps: 5
 Percent Similarity: 75.253 Percent Identity: 43.687

alignment_block:
 US-09-775-693-1 x ASSY_BUCAI ..

Align seg 1/1 to: ASSY_BUCAI from: 1 to: 403

```

19 GTGGTTCGGCTACAGTGGGGCGCTGACACCTCGTGCATCCCTGCTG 68
   |||||||
9  ValValLeuAlaIyrSerGlyGlyLeuAspThrSerAlaIleIleProtr 25
   |||||||
69 GCTGAGGAGACA...GGCTATGACGTCAATTCCTATCTGCCACATG 115
   |||||||
25 pleuysgluasnlyrasnphgIuValAlaPheValAlaAspIleG 42
   |||||||
116 GCACG...AAGAAAGACTTCGAGGAGAGCAGAGACGACCTGAACTT 162
   |||||||
42 LysIleSerIysAspLeuAsnGlyIleGluIlySlySerLeuGluSer 58
   |||||||
163 GGGGCCAAAAGGTCTTCATTGAGATCAGACGAGGAGCTTTGGAGGA 212
   |||||||
59 GlyAlaSerSerCysHisValPheAspLeuSlyGluGluPheIleGlu 75
   |||||||
213 GTTCATCTGGCCGGCCGCTCAGTCCAGCGACGACTGATGAGACCGTAC 262
   |||||||
75 nlyrValIyrProValLeuIyThrIyAlaLeuIyrgluIySerIyrl 92
   |||||||
263 TCTGCGCACCTCTCTTGGCAGGCGCTGATGCGCCGCAACAACTGGAA 312
   |||||||
92 euleuGlyThrAlaMetAlaIarProIleIleAlaIySlySgluValGlu 108
   |||||||
313 ATCCGCCAGCGGGAGGGGCGCAAGTATGTCCACAGCGCCACAGGAA 362
   |||||||
109 LeuAlaLeuAsnIleGlyAlaAsnSerLeuGlyHisGlyAlaThrGly 125
   |||||||
363 GGGGACATGATGAGTCCGCTTGGAGCTCAGCTCTACTCAGTCCGCC 412
   |||||||
125 sGlyAsnAspGluValArpPheGluMetAlaIyAlaAlaLeuAlaPro 142
   |||||||
413 AGATAAGAGTCAATTCCTCCCGGAGGATGCTGATCTCAACCGCTTC 462
   |||||||
142 snLeuAsnValIleAlaIarProIytrp....GluTrp....AsnLeu 154
   |||||||
463 AAGGGCCGCAATGACCTGATGAGTACGCAAGACACAGCGATGCCAT 512
   |||||||
155 AsnSerArgIleSerLeuLeuIySlyAspIySlyAsnIleSerTh 171
   |||||||
513 CCGGTCATCTCCCAAGAACCCCTGGAGCATGAGAGACCTCAATGACA 562
   |||||||
171 rThrAlaThrLeuGluIySlyIleIySerIyAspIyAsnSerIyPheI 188
   |||||||
563 TCGAGTACGAGGCTGATCTCGAGAACCCCAAGACCAAGCGCTTCA 612
   |||||||
188 IeSerThrIyGluIyLeuLeuGluAsnProIyPheAsnIleSerAsnGlu 204
   |||||||
613 GGTCTTACACGAAAGACCCAGGACCCAGCAAGAACACACACCTGA 662
   |||||||
205 AspCysIySerIyPheIyAlaAsnProGluAspAlaIyProGluIyProGlu 221
   |||||||
663 CATTCGAGATCGAGTTCAAAAGAGGCTCCTGTGAAGTGCACACAG 712
   |||||||

```

```

221 uTyValSerLeuGluIleuIySgluIyCysValValSerValAsnAsnG 238
   |||
713 TCAAGATGGCACCACCCACACAGACTCTCTGAGCTCTTCATGTACTCG 762
   |||
238 IuIyS.....LeuAsnProIleuIySgluValGluIleu 249
   |||
763 AACGAAGTCGGCGGAGCAAGATGCGCTGGCCGCTATGACATGTGGAGAA 812
   |||
250 AsnSerLeuGluAlaIySlyHisGlyIleGlyArpIleAspIleIleGlu 266
   |||
813 CCGCTTCAATTTGATGATGCCGAGGATATCTACGAGACCCACAGACGA 862
   |||
266 nArgLeuIleGlyLeuIySerArpGlyCysTyrgluThrProGlyIyT 283
   |||
863 CCATCTTTCACCATCTCATTTAGACATFCAGGCTTCACCATGAGCCG 912
   |||
283 hIleIleIleMetThrAlaIleIySAlaIleIyGluIleuValIleuAspArp 299
   |||
913 GAAGTCGCAAAATCAACAGCGCTGGCTTGAATTTGCTGAGCTGT 962
   |||
300 GluSerPheArpIyPheIyGluIySleIyLeuGluMetSerIleIy 316
   |||
963 GTATACGGCTTACGCGCTAGCCCTGACTGATTTGTCGCCACATGCA 1012
   |||
316 IlyrAspGlyArpIyPheSerProIleArpIySerLeuIleAlaIa 333
   |||
1013 TCGCCAAATGCCAGAGCGAGTGAAGGAAAGTGCAGGTGCTGCTC 1062
   |||
333 IAspSerLeuSerLeuGluIleIyThrGlyIyValIleuIySleuIy 349
   |||
1063 AAGGCGCAGGTGATCAATCTCGCGCGGAGTCCCATCTCTCTATACA 1112
   |||
350 LysGlySerValThrAlaValGluIySlySerProAsnSerLeuIySe 366
   |||
1113 TGAGACGTGTGAGTGCATGAACTGCAGGATGATATGAGCCAACTGATG 1162
   |||
366 rGluGluIyAlaThrPheGlyGluAspIyValIyIyIySgluIyAspA 383
   |||
1163 CCACCGGTTTCATCAACATCAATTCCTCAGGCTGAG 1200
   |||
383 IAspGlyPheIleArpLeuPheSerLeuSerIyS 395
   |||
seq_name: SwissProt_39:ASSY_AOUAE
seq_documentation_block:
ID ASSY_AOUAE STANDARD; PRT; 401 AA.
AC 067213:
DT 15-JUN-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
DE LIGASE).
GN ARGG OR AO_1140.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=VF5.
RX MEDLINE=9819666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aubay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: PENULTIMATE STEP IN ARGININE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

alignment_scores:

Quality:	800.50	Length:	404
Ratio:	2.741	Gaps:	3
Percent Similarity:	72.277	Percent Identity:	41.832

alignment_block:
US-09-775-693-1

05-113-093-1 X ASSY_AQUAE

CO: ASSY_AQUAE from: 1 to: 401

[illegible]

```

182  lytaSerIleGluLysGlyProLeuGluAspProTrpGlnGluProPro 198
610  CCAGGCTCTACAGCAAGACCAGGACCCAGGCAAGAGCCCAACACCC 659
199  GlnAspAlaTrpGlnIleTrpGlnSerProGlnAlaProAspGluP 215
660  TGACATTCGAGATGCGAGTTCGAAAAAGGGTCCCTGTGAAGTACCA 709
215  oGluTrpValThrValGlyPheGluLysGlyLysProValTyrLeu... 230
710  ACGTCAGGATGGACCCACCACAGACCTCTTGGAGCTTCACATGTC 759
231  .....AsnGlyGluArgTyrGluGluGlnTrpLysLeuIleAlaAsn 244
760  CTGACAGAGTCCGGGCGAAGCATGGCGGTGGCCGATTTACATCGTGA 809
245  LeuAsnGlnIleAlaGlyArgHisGlyValGlyArgIleAspMetValG 261
810  GAACCGTTCATTTGAAATGAAGTCCGAGTATCTACGAGAGCCAGAG 859
261  uAsnArgLeuValGlyIleLysSerArgIleTyrGlnAlaProGlyA 278
860  GCACCATCTTACCATGCTCATTTAGACATCGAGAGCTTACCATGAGC 909
278  laMetValLeuTrpGlnAlaTrpArgAspLeuSerLeuValIleAsp 924
910  CGG...GAAGTCCGCAAAATTCAAACAAGGCTGGGCTTGAATTTGCTG 956
295  ArgPheThrPheHisTyrPheLeuThrHisIleProHisGlnTyrAla 311
957  GCTGCTGATACCGGTTTACGGCCCTGAGCCCTGAGTCAATTTGTCCGC 1006
311  sheuValTyrGluGlyLeuTrpPheThrProLeuArgGlnAlaLeuAsp 328
1007  ACTGATCGCCAAATCCCAAGGAGCGAGTGAAGGAAAGTGCAGTGTGC 1056
328  laPheThrAsnLysIleAlaGluPheAlaTrpGlyGluValArgLeuLys 344
1057  GTTCCTCAAGGGCCAGGTGTACATCTCCGCGGGAGAGCCCACTGCTCT 1106
345  LeuTyrLysGlySerValSerValGlyArgTrpSerProAsnSerIle 361
1107  CTACAAATGAGAGCTGTGAGCATGAAACGTGCAGGGTATATGAGCCCA 1156
361  uTyrValGluGluLeuAlaIleThrTyrSerGluLysAspGlnPheAspGln 378
1157  CTGATGCCACCGGTTTCATCAACATCAATCAATCCCTCAGGTGGAAGAATAT 1206
378  leaIleGlyLysHisPheThrLysValTrpLysLeuProLeuLysValLeu 394
1207  CATCGCTCTCCAG 1218
395  GLyArgValArg 398

seq_name: SwissProt_39:ASST_LACIA

seq_documentation_block:
ID      ASST_LACIA      STANDARD:      PRT:      398 AA.
AC      P57799.
DT      20-AUG-2001 (Rel. 40, Created)
DT      20-AUG-2001 (Rel. 40, last sequence update)
DT      20-AUG-2001 (Rel. 40, last annotation update)
DE      ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
GN      LACIAS).
OS      Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae.
OX      NCBI_TaxID=1360;
RN      [1]
RS      SEQUENCE FROM N.A.
RC      STRAIN=TL1403.
XX      MEDLINE=21235186; PubMed=11337471;

```

```

134   erleAlaIaProGluIleuAlaValIleAlaProValArg.....GluTrp 148
451   TACAACCGGTTCCAGGCGCGCATGACCTGATGAGAGTACGAAACACACA 500
149   His.....TrpAlaArgGluGluGluGluGluTrpAlaAsnGlns 162
501   CGGATTCCTCCATCCCGGTCATCTCCACAGAACCCGGTGAGCATGATGAGA 550
162   nclYAlaProIleProAlaAspLeuAspAsnProTrpYserIleAspMetA 179
551   ACCCTATGCGACATCAGCTTACGAGGCTGGAATCTGTGAGAACCCCAAGAAC 600
179   snleuTrpGlyArGAlaIleGluAlaGlyValLeuGluAsnProTrpAsn 195
601   CAAGCCCTCCAGGCTCTACACAGAACCCGACAGCCCAAGACCCC 650
196   ThrCysProGluAspAlaIlePheMetTrpAsnSerValGluAsnLapR 212
651   CAACACCCCGCATTCCTGAGATGACGATTCAAAAAAGGGTCCCTGTGA 700
212   oAsnGluAlaGluIlePheIleGluValGluPheGlyGluGlyLeuProIleA 229
701   AGGTGACCAACGTCAGAGATGCGACCCACACAGACCTCTTGAGACTC 750
229   lAlaLeu...AsnGlySerLeuGluLeuHis.....GluIle 240
751   TTCATGTACCTGAACGAAAGTCGCGGCGCAACATGAGCTGGGCGCTATTGA 800
241   IleGlySerValAsnIleIleAlaGlyLysHisGlyIleGlyArgIleAs 257
801   CATGTGTGAGAACCGCTTCATTGGAATGAAGTCCCGAGATTCACAGAGA 850
257   PHisIleGluAsnArgLeuValGlyIleLysSerArgGluPheTrpGluC 274
851   CCCGACGAGCGACACATCCTTTACCATGGCTCATTTAGACATGAGAGCCTTC 900
274   ySProAlaAlaIleThrLeuLeuLysAlaHisLysAspLeuGluAsnLeu 290
901   ACCATGTGACCGGGAAGTCCGCAAAATCAACAAAGGCTGGGCTTGAATT 950
291   ThrPheValArgGluLeuAlaHisPheLysProValIleGluAsnGluIle 307
951   TGTGAGCTGTGTATACCGGTTTACGGCCACAGCCCTAGTGTGAAATT 1000
307   uAlaAsnLeuIleTrpAsnGlyLeuTrpPheAsnProAlaThrLysAlaL 324
1001   TCCCGCGCATGCGCCCAAGTCCCGAGAGCGAGTGAAGGAAAGTCGAG 1050
324   euIleAlaTrpLeuAspGluTrpGlnLysValValAsnGlyIleValLys 340
1051   GTGTCCGCTCCTCAGGCGCAGGTGATACCTCTCGGCGCGGAGATGCCCAT 1100
341   IleLysLeuTrpGlyGlyLeuAlaThrProValGlyArgLysSerThrAs 357
1101   GTCCTCTCTACATGAGAGCATGCTGTCGATGACGATGACGAGGATTAAG 1150
357   nSerLeuTrpSerGluLysLeuAlaLehTrpThrAlaAlaAspGluPhea 374
1151   AGCCAACTGATGCCACCGGCTTCATCAATC 1182
374   spoJnAlaIleAlaValAlaGlyIleLysLeu 384
seq_name: swissProt_39:ASSY_SYNY3
seq_documentation_block:
ID ASSY_SYNY3 STANDARD; PRT; 400 AA.
AC P77973;
DI 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
LIGASE).
GN ARGG OR SLR0585.

```

OS *Synechocystis* sp. (strain PCC 6803).
 RX Bacteria: Cyanobacteria: Chroococcales: *Synechocystis*.
 ON NOBL_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97061231. PubMed=8905231;
 RX Kaneo T., Sato S., Kotani H., Tanaka A., Asanuma E., Nakamura Y.,
 RA Miyajima N., Hisosawa M., Sugiyama M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nairo K., Okumura S.,
 RA Shiba S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Rep. 3:109-136(1996).
 CC -I- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARATE = AMP +
 CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
 CC -I- PATHWAY: PENULTIMATE STEP IN ARGININE BIOSYNTHESIS.
 CC -I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D90917; BAA18641.1; -
 DR InterPro: IPR001518; Arginosuc_synth.
 DR Pfam: PF00764; Arginosuc_synth.1.
 DR ProDom: PD003544; Arginosuc_synth.1.
 DR PROSITE: PS00564; ARGININOSUCCIN_SYN.1; 1.
 DR PROSITE: PS00565; ARGININOSUCCIN_SYN.2; 1.
 KW Arginine biosynthesis; Liasase; AMP-binding; Complete proteome.
 SQ SEQUENCE 400 AA; 44485 MW; 37BDB736CBB82C05 CRC64;

```

alignment_scores:      Quality: 764.00      Length: 400
                       Ratio: 2.709      Gaps: 6
Percent Similarity:    69.458      Percent Identity: 41.626

alignment_block:
US-09-775-693-1 x ASSY-SYNY3      ..

Align seg 1/1 to: ASSY-SYNY3 from: 1 to: 400

```

[illegible]

295 GCGCGCAACGACGAGGAAATCGGCGGACGGGAGGGGCCAGATGTGTCTC 344
 101 AAlaymetleuValGluAlaAlaGluLysYrGlyAlaAspAlaValAl 117
 345 CGACGGCGCCACAGAAAGGGAGGACATAGGTCCGGTTTGAGGTACAGCT 394
 117 aHISgLIcysThrGlyLysgLIyAsnAspIValAlaArgHepHisLeSerI 134
 395 GCTACTACAGGGGCGCCGAGATTAAGGTGATTTGGTCCC.....TGG 435
 134 LemetAlaLeuAsnProHisLeuLysValLeuAlaProAlaArgGluTyr 150
 436 AGGATGCGTCGAATTCTTCAACCGGTTCACAAAGGGCGCCAGTATGACG 485
 151 LysMet.....SerArgGluGluThrIleAl 159
 486 GTTCGGCAAGGCAACAGGGATTTCCATCCGGGTACTCTCCAGAACCCGT 535
 159 aTYrGlyGluAlaArgYrGlyValGluSerProValLysLysSerProI 176
 536 GGAGCATGATGATGAACCCATCATGATGCATCTACCTACGAGGCTGGAATCC 585
 176 YrSerIleAspArgAsnIleLeuGlyArgSerIleGluAlaGlyProLeu 192
 586 GAACACCCCAACAAACCAAGCGCTCCAGGCTCTTACAGCAAGACCGAGA 635
 193 GluAspProMetThrGluProThrGluGluIleYrLeuMetThrLysAl 209
 636 CCGACGCAAGACCCCGCCACACCCCTGACATTTCTTCGAGATTCGAGTTAA 685
 209 aIleAlaAspThrProAspGluProGluTyrValAspIleGlyProGluL 226
 686 AAGGGTCCCTGTGAAGGTGACCAAGCTCAAG...GATGGACACCCAC 732
 226 ysgLIleProValSerLeuAsnGlyAlaMetLeuAspProValThr... 241
 733 CAGACCTCCTTGAGCTCTTCATGTACCTGTGAACAGAGTCCGGGCAAGA 782
 242LeuValGluAlaGlyLeuAsnGluIleAlaGlyAsnHI 253
 783 TGGCTGGGCGGATTGATCATGCTGGAGAACCCCTTCAATGGAAATGAGT 832
 253 scLIyAlaGlyArgLeuAspMetValGluAsnArgValaValGlyIleLys 270
 833 CCGGAGTATCTACGACCCCGACAGCAGACCATCTCTTACCAGTCTCAT 882
 270 eArgGluIleYrGluAlaProAlaLeuIleValIleuIleAspAlaHis 286
 883 TTAGACATCGAGGCTTCAACATGAGCCGGGAAGTGGCCAAATCAACA 932
 287 ArgAspLeuGluSerLeuThrGlnThrAlaAspValIleuHisYrLysAs 303
 933 AGGCGTGGCTTGAATTTGCTGAGCTGGTGTATACCGGTTTACGGCCTA 982
 303 nThrValGluGluIleYrSerGlnLeuIleYrArgGlyLeuThrPyrLys 320
 983 GCGCTGAGTGAATTAATTTGGCCCGCCAGCATGATGCCAATCCCGAGAGCCA 1032
 320 eProLeuLysGluAlaLeuAspAlaIleIleValYrThrGlnGluArg 336
 1033 GTGAAGGCAAACTGAGGTGCTCCGCTCTCAAGGCGCAGGTGATCATCTC 1082
 337 ValThrGlyMetValaValArgLysPheHepLysGlyAsnAlaValAl 353
 1083 CGCGCGGAGTCCCACTGTCTCTCTACATGAGGAGCTGTGAGCATGA 1132
 353 aGlyAlaGlyLysSerAspYrSerIleYrAspAlaGluLeuAlaThrTyrG 370
 1133 ACGTGCAGGAGTATATATAGCAACTGATGCAACCGGGTTCATCAACATC 1182
 370 YMetGluAspIlePheAspHisLysAlaAlaGluGlyPheIleYrIle 386
 1183 AATTCCTTCAGGCTGAG 1200

387 trpGlyLeuProThrLys 392

seq_name: SwissProt_39:ASSY_ARCFU

seq_documentation_block: STANDARD; PRT; 390 AA.

AC 028032: 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE-ASPARTATE
 DE LIGASE).
 CN ARGV OR AF2252.
 OS Archaeoglobus fulgidus.
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 CC Archaeoglobus.
 CC NCBI_TaxID=2234;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RC MEDLINE=98049343; PubMed=9389475;
 RX Kleink H.-P., Clayton R.A., Tomb J.F., White O., Nelson K.E.,
 RA Richardson K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Fleischmann D.L., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirschner S., Ketch C.I., McNeil L.R., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne T., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., DAndrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RA "the complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
 CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- PATHWAY: PENULTIMATE STEP IN ARGININE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE000949; AAB89005.1;
 DR trfGR: AF2252;
 DR InterPro: IPR001518; Arginosuc_synth.
 DR Pfam: PF00764; Arginosuc_synth. 1;
 DR ProDom: PD003544; Arginosuc_synth. 1;
 DR PROSITE: PS00564; ARGININOSUCCIN_SYN_1; 1;
 DR PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1;
 DR Arginine biosynthesis: ligase: ATP-binding. Complete proteome.
 KW ARGinine biosynthesis; ligase; ATP-binding; Complete proteome.
 SEQUENCE 390 AA; 44104 MW; 31334D84F696EBD CRC64;

alignment_scores: Quality: 750.50 Length: 407
 Ratio: 2.544 Gaps: 12
 Percent Similarity: 72.482 Percent Identity: 43.735

alignment_block: 1 x ASSY_ARCFU

Align seg 1/1 to: ASSY_ARCFU from: 1 to: 390

19 GTGGTCTGGCTACAGTGGCGCTGACACCTGTCGATCCGCTGTG 68
 |||||||

3 ValValLeuSerTyrSerGlyGlyLeuAspThrThrValCysIleProLe 19
 69 GCTGAAGAGCAAA...GGCTATGAC...GTCAATGGCTATGTGGCCACA 112
 |||||||
 19 uLeuysGluLysTyrGlyPheAspGluValIleThrValThrValAspI 36
 113 TTGGCCACAGAGAA...GACTTCAGAGAACCCAGAGAACGACTGTAG 159
 |||||||
 36 IeGlyGluProGluAlaAspIleLysGluAlaGluGluArgGlyLysIys 52
 160 CTGGGGGCCAAAAGGTGCTTCATGAGATGCTCAGCGAGGAGTTTGTGA 209
 ||| |||
 53 Tyr...AlaAspLysHisTyrThrIleAspAlaLysGluPheValas 68
 210 GGAGTTCATCTGGCGCGCCATCCATCCAGCGCACCTGATGAGAGCCGCT 259
 |||||
 68 pSer...LeuPheMetLeuLysAlaAsnGlyAsnGlyIleLysIle...T 83
 260 ACCTCTGGCGACCTCTCTTCCGACAGCGCTGCATCGCCGCAACAGTG 309
 |||||
 83 yValLeuGlyThrAlaLeuAlaArgProLeuIleAlaGluLysValI 99
 310 GAATCGCGCGCGAGGAGGCGCAAGTATGTGCCACGCGGCCACAG 359
 |||||
 100 GluValAlaLysLysGluGluAlaGluAlaIleHisGlyCysThrG 116
 360 AAAGGGAGACGATCAGTCCGCTTGGCTGACCTCTACTGCTGCTGCCC 409
 |||||
 116 yLysGlyAsnAspGluLeuArgPheGlu...AsnIlePheA 129
 410 CCGAG...ATAAGGTCAATGCTCCCTGGAGATGCTGATATCTAC 453
 |||||
 129 rGluHisGlyPheLysValIleAlaProValArgGluLeuAsnLeuThr 145
 454 ACCGGTTCAGAGCGCGCAATGATGATGATGATGATGATGATGATGATG 503
 |||||
 146 ArgGluIleProLeu...IleGluIleAlaArgGluHisG 157
 504 GATTCCTCCGCTGCTGATCCAGAACCCGCGAGATGATGATGATGATG 553
 |||||
 157 yIleGluValProAlaThrLysGluLysProIleSerIleAspGluAsn 174
 554 TCATGACATCAGCTGATGATGATGATGATGATGATGATGATGATGATG 603
 |||||
 174 eutPTrSerArgSerValGluGlyLysLeuGluAspProSerPheGlu 190
 604 GCGCTCCAGCTCTACACAGAACCCAGAACCCAGAACCCAGAACCC 653
 |||||
 191 ProProGluAspIleTyrGluTyrPheIleSerProGluLysAlaPro 207
 654 CACCCGTGACATTCGAGATCGATGATGATGATGATGATGATGATGATG 703
 |||||
 207 pLysProGluIleValLysIleAspPheGluLysGlyValProValAla 224
 704 TGACCAACGTCAAGATGCGACACCCAGACCTGCTGAGCTCTTC 753
 |||||
 224 euAsnAspGluArgMetGly...GlyPheGluLeuIle 235
 754 ATGATCTGACAGCAAGTCGCGGCGCAAGCATGCGCGGCGCTGATGAT 803
 |||||
 236 LysAlaLeuAsnGluIleGlyLysHisGlyValGlyArgThrAspMe 252
 804 CGTGAGACACCGCTTCATGTAAGTCCGAGATGATGATGATGATGATG 269
 |||||
 252 tIleGluAspArgValLeuGlyLeuLysAlaArgGluAsnTyrGluHis 293
 854 CAGCAGCAGCATCTTACATGCTCATTTAGCATGATGATGATGATGATG 303
 |||||
 269 roAlaAlaThrIleLeuIleThrAlaHisArgAspLeuGluAsnLeuVal 285
 904 ATGAGACCGGAGAGTGCACAAATCAACAGAGCTGCGCTGAAATTCG 353
 |||||
 286 LeuSerArgArgGluLeuLysPheLysPheValGluGluIleProAl 302

seq_name: SwissProt_39:ASSY_METJA

DR Pfam; PF00764; Arginosuc_synth.1.

alignment_scores:

ASSY_METJPA

22 GTTCTGGCCCTACAGTCCCCCCCCCC-

72 GAACCAACAA
cyscysleuLysLeu 22

110 39

...GluGluIleLysGluValGluGluLysAlaLysLysLeuIv 55

50 VALLEULYSHISTYrThrIleaspAlaLYSGluGlnpheValIrua... ..

72 rlepheargalal1elYSA1aasn1amotT...

```

88 eUserThraJalaLeuJAarDroT      |||||:::||| |||||

```

[illegible]

|||||
|||..

...AGCTGCTACTGGCCCCCAGA 415

...TCCCTGGAGGATGCCCTGAATTCTACCAACCGGTCAAG 465

58CCGCAATGACCTGATGGAGTACGCCAAGCAACACACGGGATTCCCATCCGCTT

GGTCACTCCCAAGAACCGTGAGCATGGATGAGAACCCTTC

polylys 182

566 GCTACGAGGCTGGAATCCTGGACAACCGCAATTTC

[illegible]

ASPLYSGlUGlUGlUl 214

-----GlyValProValAlaIleasnGlyGlu 231

```
716 AGGATGGACACCCAGACCTCTTGAGCTCTTCATGTACCTGAC 765
||
||
231 ys.....LeuGluProValGluLeuIleLysLysAlaasn 242
||
766 GAAGTCGGCGGCAAGCATGGCGCTGGCCGTATGACATGTGGAGAACCG 815
|||||
243 GluIleAlaGlyLysHisGlyValGlyArgIleAspIleIleGluAspAr 259
||
816 CTTCATTGGATGAAGTCCCGAGGTATCTACGAGACCCAGCAGGACCA 865
||
259 gIleIleGlyLeuLysSerArgGluAspTyrGlyCysProGlyAlaValL 276
||
866 TCCTTTACCATGCTCATTTAGACATCGAGCCCTCACCATGACCGGAA 915
||
276 euleuLeuThrAlaHisLysAlaLeuGluGluLeuValLeuThrArgAsp 292
||
916 CTCGCGAAATCAACACAGGCTGGGCTTGAATTTGCTGAGCTGGTGA 965
||
293 GluLeuArgPheLysGluIleValAspSerLeuTyrGlyGluLeuIleTy 309
||
966 TACCGGTTTACGGCCTAGCCCTGAGTGTGAATTTGTCCGCCACTGCATCG 1015
||
309 rLysGlyLeuTyrPheAspProLeuArgGluAspLeuAspAlaPheIleA 326
||
1016 CCAAGTCCCGAGGACCGAGTGGAGGAAAGTGCAGGTGCCGTCTCAAG 1065
||
326 sPluYstrGlnGlnArgValThrGlyThrValLysValLysLeuPheGly 342
||
1066 GGCAGGTGTACATCTCGGCGCGGAGTCCCACTGCTCTTACAAATGA 1115
||
343 GlyThrAlaArgValValGlyArgAspSerProTyrAlaLeuTyrSerLy 359
||
1116 GGAGCTGTGAGCATGACGACGTGCAGGSGTATGAGCCAACTGATGCCA 1165
||
359 sGluLeuValSerPheAsp...GluLysGluIleAspGlnLysGluLeuA 375
||
1166 CCGGTTTCATCAACATCAATTCCTCAGGCTGAAGAAATATCATCGTCTC 1215
||
375 laGlyMetVal.....LysTyrHisGlyLeu 383
||
1216 CAGAGC 1221
||
384 GlnAla 385
```



```

160 CTGGGGCCAAAGGTTTCATTGAGATGTCAGACGAGGAGTTTGCA 209
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 ThcIyAlaserlyvalylvalgluasprleuargcluhelvalth 68
210 GGAGTTCATCTGGCCGCCATCCAGTCAGCCAGCATGTATGAGACCGCT 259
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
68 rAspYrIlePheThrAlaLeuLeuGlyasnAlaMetCylrgluyargt 85
260 ACCCTCGGCGACCTCTTTCAGGCGCTGATCGCCCGCAACAAGTG 309
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85 yrlleuGlyThrAlaIleAlaargProleuIleAlaLysargcluhel 101
310 GAATGCCCGCAGCGGAGGCGCAAGTGTGTCCAGCGCGCACAG 359
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 GluhlelAgIluysgluclyAlaGlnYrAlaAlaHlsIylalathrl 118
360 AAAGGGAGCATCAGTCGCGGTTGAGCTCAGCTGCTACTGCGGCC 409
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
118 yLysGIYAsnAspGlnValaArgPheGluLeuThrYrAlaAlaLeuAsp 135
410 CCCAGTAAAGTCAATGTCCTCCGAGGAGATGCTGAATTCATACACCG 459
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
135 roAsnleuYsValIleSerProTrpIlyAspProGluPheLeuAlaLys 151
460 TTCAGGCGCGCAATGACCTGATGAGTACGCAACAGCAACCGGATTC 509
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 PheYsGIYArThrAspleuIleAsnYrAlaMetGluysgllylePr 168
510 CATCCGCGTACCTCCAGAACCCGTCGAGCATGATGATGATGATGAT 559
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 oIleYsValserlyslYsArgProTrpSerGluAspGluAsnLeuMet 185
560 ACATCAGTACGAGCTGGAATCTTCGAGAACCCAGCAACCAACGCGCT 609
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
185 IsIleSerHlsIalaglyLysLeuGluAspProAlaHlsIleProasp 201
610 CCAGTCTCTACGAAACCCAGCGCCAGCAACGCGCCACACCGCC 659
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
202 GluaspyAlPheThrYrThrYrAlaSerProIlyAspAlaProaspGlu 218
660 TGACATTCTGAGATGAGTCAAAAAGGCGCTCTGATGATGATGATGAT 251
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 urhIleuLeuGluIleHlsIlePheGluAsnGlyIleProvalYsValAla 235
710 ACCTCAGAGTGCACACCCAGCAACGCTTGTGAGCTTTCATGTAT 251
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
235 snleuYAspGlyThrGluYsThrAspProleuGluLeuPheGluYr 251
760 CTGACGAAAGTCGCGGCAAGCATGCGCGGCGCTATTGACATGATGAG 809
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
252 leuAsnGluValGlyAlaLysAsnGlyValGlyArgLeuAspMetValG 268
810 GAACCGCTTCAATGAAATGAGTCCCGAGTATCTAGACAGCCGACAG 859
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
268 urAsnArgPheIleGlyIleYsSerArgGlyValYlGluThrProGlyA 285
860 GCACCATCTTACCATGCTATTGATGATGAGGCTGATGATGATGATG 909
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
285 lathrlleuThrIleAlaHlsIleArgAspLeuGluYlIleHlsIle 910
910 CGGGAAGCGCCAAATCAAAAGAGCGCTGATGAAATTTGCTGAGACT 959
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
302 LysGluValMetHlsIleuArgAspMetLeuAlaProLysPheAlaGlu 318
960 GGTGATACCGGTTTACGCGCTGATGATGATGATGATGATGATGAT 1009
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318 uIleYrYsAsnGlyPheTrpPheSerProGluMetGluPheLeuAla 1010
1010 GCATCGCAAGTCCAGGAGCGAGTGAAGGAAAGGAGGAGGCTCGCTG 1059
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
335 lathrlYsAlaGlnleuAsnValThrGlyLysValThrValserIle 351
1060 CTCGAAGCGCAGGTCTACATCTCGCGGAGACTCCCACTGCTCTCTA 1109

```

```

352 TyrlYsGlyAsnValMetProvalAlaArgYsSerProTrpIlySer 368
1110 CAATGAGGAGCGTGTGACATGAACTGACGAGGCTGATTTATGACCACTG 1159
368 rAsnProGluLeuSerSerMetAspValGluGlyPheAspAlaThrA 385
1160 ATGCCACCGCGTTCATCAATCATCAATTCCTAGGCTCAAGATATCAT 1209
385 spserLysGlyPheIleAsnIleHlsIleAlaLeuArgLeuYs...ValHls 400
1210 CGTCTCAGAGCAAG 1224
401 GluhleuValLysLys 405

seq_name: sp_bacteria:Q9RMJ4
seq_documentation_block:
ID Q9RMJ4 PRELIMINARY; PRJ: 402 AA.
AC Q9RMJ4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
DE LIGASE).
OS Delinococcus radiodurans.
OC Bacteria; Thermus/Delinococcus group; Delinococcales; Delinococcus.
OX NCBI_TaxID=1299;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=RI.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Makarova K.S., Aravind L., McDonald L., Uitterback T., Zaslowski C.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Delinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
CC -I- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -I- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC -I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -I- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
DR EMBL: AEO01924; AAF10250.1;
DR TIGR: DK0674;
DR InterPro: IPR001518; Arginosec_synth.
DR Pfam: PF00764; Arginosec_synth.
DR ProDom: PD003544; Arginosec_synth; 1.
DR PROSITE: PS00564; ARGININOSUCCIN_SYN_1; 1.
DR PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1.
KW ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;
KW Urea cycle.
SQ
SEQUENCE 402 AA; 44877 MW; 8694AD1424A99835 CRC64;

alignment_scores:
Quality: 1034.00 Length: 411
Percent Similarity: 76.886 Gaps: 6
alignment_block:
US-09-775-693-1 x Q9RMJ4
US-09-775-693-1 x Q9RMJ4
Align seg 1/1 to: Q9RMJ4 from: 1 to: 402
7 ACCAAGGCTCGGTGCTGACCTAGAGGCGCGCTGAGACCTCGTG 56
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

2 SerLysGIuLysIleValIleuAlaTyrSerLysGlyLeuAspThrSer11 18
57 CATCTCGTGTGCTGAAG...GAACAAGCTATGACGTCAATTGCTATC 103
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 eileLeuLysTrpLeuGlnThrGluArgAsnLysPheValValLysPheT 35
104 TGCCCAACATTTGGCCAGAGAAAGACTTCGAGAGCCAGCAGAGAAAGCA 153
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 hrLAlaSpLeuGlnGlnLysAspGluValGlnGlnAlaIleValLysAla 51
154 CTGACCTGGGGGCCAAAGGTGTTCAATTGAGATGTCAGCAGGAGACTT 203
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 LeuAsnThrGlyAlaValAlaIleAlaTyrAlaLeuAspLeuArgGlnLuph 68
204 TGTGAGAGATTCATCTGGCCGCCATCCAGTCCAGCCGACCTGTATGAG 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 eValArgAspLysPheValPheProMetMetArgSerSerAlaLeuTyrGlnG 85
254 ACCGCTACCTCCGCGGACCTCTCTTCGACGCGCCCTGATCCGCGCAAA 303
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 LysTyrTyrLeuLeuGlnLysSerIleAlaArgProLeuIleAlaLysLys 101
102 MetValGlnIleAlaGlnLysGlnLysGlnLysAlaValAlaIleSerHisGlyAl 118
304 CAAGTGAATTCGCCCGCAGCGGAGGGGCCAAGTATGTCTCCAGCGCGC 353
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 MetValGlnIleAlaGlnLysGlnLysGlnLysAlaValAlaIleSerHisGlyAl 118
354 CACAGGAAAGGGGAGACATCATGCTGGGTTTGAGCTCAGCTGCTACTGAC 403
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 aThrGlyLysGlyAsnAspGlnValArgPheGlnMetSerAlaTyrAla 135
404 TGGCCCCCAGATPAAAGTCATGCTCCCTCGAGGATGCTCTGAATTTCTAC 453
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 eulysProAspIleValThrValAlaProThrArgAspTrpasp..... 149
454 AACCGTTTCAAGGGCGCAATGACCTGATGAGTAGCCAAAGCAACAGCG 503
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 .....PheGlnGlyArgAlaAspLeuGlnAlaPheAlaArgGlnHisGly 164
504 GATTCCCATCCCGGTGACCTCCCAAGAACCGGTGAGATGATGAGAACG 553
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 yIleProValProThrThrLysLysAspProThrPseMetMetAlaAsnM 181
554 TCATCAGATCATGCTACGAGGCTGGAATCTCGAAGAACCCCAAGAACCA 603
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 eLeuHisIleSerTyrGlnLysGlnLysGlnLysGlnLysGlnLysGln 197
604 GCGCTCCAGGCTCTACACGAAAGCCAGGACCCAGCAGCAAGCCCAAA 653
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 ProProThrHisMetPheLysLeuThrValAlaAsnProGluAspAlaProse 214
654 CACCCCTGACATTCGAGATCGAGTTCAAAAAGGGGTCCCTGTCGAGG 703
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 tGluAlaGlnLysValGlnLysGlnLysValAlaAsnLysAspProValSer1 231
704 TGACCAACGTCAGAGATGGACACCCAGACCTCTTGAGACTCTC 753
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 leAsnGly.....GluGlnLeuSerProAlaAlaLeuLeu 242
754 ATGTACTCTAAGAGAGTGGCGGCAAGAGATGGCGGCTATTTGACAT 803
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
259 uValGlnAsnArgPheValGlyMetLysSerArgGlyValTyrGlnThrP 276
854 CAGCAGGACACATCTTTACATGCTCATTTAGACATGAGCGCTTACG 903
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 roGlyGlyThrLeuLeuTyrHisAlaArgAlaValGlnSerLeuThr 292
904 ATGACCGCGGAAGTGGCAAAATCAACAAAGCGCTGGCGTTGAAATTCG 953
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
293 LeuAspArgGlnValLeuHisGlnArgAspAlaLeuGlyProLysTyrAl 309

```

```

954 TGAGCTGTGTATACCGGTTATACGGCTTACCGCTGAGTGTGA..... 996
309 agLLeuValLysAsnGlyPheThrPheAlaProGlnArgGlnAlaLeuG 326
997 .....TTTCTCCGCCACATCCATCCAGTCCAGGACGAGCAGTGAAGG 1041
326 InValTyrPheAspHis...ValAlaLysSer.....ValThrGly 338
1042 AAGTGCAGTGTCCGCTCTCAAGGGCCAGGTGTACATCTCTGGCGGGA 1091
339 ThrAlaArgLeuLysLeuTyrLysGlnLysCysIleValAlaGlyArg 355
1092 GTCCCGCATCTCTCTTACATGAGAGAGCTGTGACATGAACGTGACG 1141
355 sAlaGlnArgSerLeuTyrAspLysAspLeuValSerPheGlnAlaGlyG 372
1142 GTGATTATGAGCCCACTGATGCCACCGGCTTCATCAACATCAATTCCT 1191
372 LysAspTyrAsnGlnHisAspAlaGlyAlaPheIleLysLeuAsnSerLeu 388
1192 AGGCTGAAGAAATATCATCTCTCCAGAGCAAG 1224
389 ArgMetArgValGlnLysArgValGlnAspLys 399
seq_name: sp_bacteria:Q9ABU1
seq_documentation_block:
ID Q9ABU1 PRELIMINARY; PRT; 408 AA.
AC Q9ABU1;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE.
GN CC0129.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RX MEDLINE=21173698; PubMed=11259647;
RA Eisen J., Heidelberg J.F., Almey M.R.K., Ohta N., Maddock J.R.,
RA Potocny J., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.C., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AB005687; MAK22116.1; -.
DR TIGR; CC0129; -.
KW Complete proteome.
SQ
SEQUENCE 408 AA; 45330 MW; 06571CHDC38B0FBD CRC64;

```

```

alignment_scores:
  quality: 948.50      length: 401
  ratio: 3.090        gaps: 8
  percent similarity: 76.559      percent identity: 50.125

```

```

alignment_block:
US-09-775-693-1 x Q9ABU1

```

```

Align seg 1/1 to: Q9ABU1 from: 1 to: 408

```

```

19 GTGGTTCTGCTACAGTGGCGCTGGACACCTGCTGATCTTCGCTG 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9 ValValIleuAlaTyrSerLysGlyLeuAspThrSerIleIleLeuLysTr 25
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69 GCTGAGG...GAACAAGCTATGAGCTGCTGCTATCTGGCAACATTC 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

25 pleuGlnThrGluTyrGluAlaGluValIleThrPheThrAlaAspLeu 42
116 GCCAAGAAAGAACTTCGAGAGACCGAGGAAGAGACGATGAGCTGGG 165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 LysGlnGluGluGluIleGluProAlaArgAlaLysAlaLeuAlaGly 58
166 GCCAA.....AAGTGTTCATTGAGATGTCAGCAGGAGTGTGGA 209
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
59 ValLysProGluAsnIlePheIleGluAspValArgGluGluPheVal 75
210 GGAATTCATCTGGCGGCATCCAGTCCAGCCGACCTGTATGAGACCGT 259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
75 AspTyrValPheProMetPheArgAlaAsnThrValTyrGluGluGln 92
260 ACCTCTGGCCACCTCTTTCGAGGCGCTGATCCGCCCAACAGAGT 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 TyrLeuGluGluThrSerIleAlaArgProLeuIleAlaLysGlnIle 108
310 GAATCGCCCAAGCGGAGGCGCAAGTATGTGCCAGCGGCCACAG 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 GluIleAlaArgLysMetGluAlaAspAlaValSerHisGluAlaThr 125
360 AAGGGAAACGATCAGTCCGGTTTGAGTCACTGCTACGACGGCC 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 LysGluAsnAspGluValArgPheGluLeuGluTyrTyrGluLeuGlu 142
410 CCCAGATAAGCTATGCTCCCTGGAGAGATCGCTGAATTCACACCG 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
142 LysPheThrValIleAlaProTyrPargGluTyrPasp..... 154
460 TTCAGGGCGCAATGATGATGAGTGAAGCAAGCAACGCGATTC 509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 PheLysSerArgGluAlaLeuLeuAspPheAlaGluLysGlnIleGlu 171
510 CATCCCGGTCATCCCAAGAAC.....CCGTGAGCATGATGAGAACC 553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 IleThrLysAspLysArgGluAlaProPheSerValAspAlaAsn 188
554 TCATGCATCAGTACGAGGCGTGAATCCGAGAACCCCAACACCA 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
188 euleHisSerSerSerIleGluLysValLeuGluAspProAlaValGlu 204
604 GCGCTCCAGTCTCTACACGAGACCCAGACCCAGACCCCAAG 653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
205 AlaProGluPheValHisMetArgThrIleAlaProGluAspAlaPro 221
654 CACCCCTGACATCTCCGAGTCCGAGTCAAAAAGGGCTCCGTGAAG 703
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 PalarThrIleIleThrIleAspPheGluLysGluAspProValAla 238
704 TGACCAACGTCAGATGAGCAGCAGCAGACGCTCTGGAGCTTC 753
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
238 Le.....AspGluValAla...MetSerProAlaThrLeu 249
754 ATGTAACGAAAGAGTCCGCGGCAAGCATGCGTGGCGGTATGACAT 803
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 ThrLysLeuAsnGluLeuGluArgAspAsnGluValGluArgLeuAsp 266
804 CGTGGAAACCGCTTCATGTGAAGTCCGAGTATCTAGACACCC 853
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 uValGluAsnArgPheValGluLysSerAlaGluValTyrGluThr 283
854 CAGCAGCAGCAGCTTACCATGATGACATGACAGCAGCAGCTTAC 903
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 LysGluGluThrIleLeuLeuAlaIleHisArgGluLysLeuSerIle 299
904 ATGAGCCGGAAGTCCGCAAAATCAACAAGCGCTGGCTGAATTC 953
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
300 LeuAspArgGluAlaMetHisLeuLysAspGluLeuMetProLysTyr 316
954 TGAGCTGTGTATACCGTTTACGCGCTCCGCTGAGTGTGAATTC 1003
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 AserLeuValLysAsnGluPheThrPheSerProGluArgGluMetLeu 333

```

```

1004 GCCATGCAATGCCAATGCCAGAGCGAGTGAAGGAATGCGAGTC 1053
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
333 LnaIleAlaIleAspTyrSerGlnAspLysValThrGluValArgVal 349
1054 TCCGCTCCAGAGGCGCAGGTGATACATCCGCGCGGAGATCCCAT 1103
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
350 LysLeuTyrLysGluAsnValThrValIleGluArgGluSerProTyr 366
1104 TCCTACATGAGAGCTGCTGACATGAACGTGACGCTGAT.....T 1147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 rLeuTyrAspGluAspLeuValThrPheGlu...GluGluLysValAla 382
1148 ATGAGCAACATGATGCCCGGTTTCATCAATCAATATCCCTGAGCT 1197
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 TyrAspHisArgAspAlaGluGlyPheIleLysLeuAsnAlaLeuArg 398
1198 AAG 1200
399 Arg 399

```

```

seq_name: sp_bacteria:Q9HY84
seq_documentation_block:
ID Q9HY84 PRELIMINARY; PRT; 405 AA.
AC Q9HY84:
DT 01-MAR-2001 (TREMBLrel, 16, Created)
DT 01-MAR-2001 (TREMBLrel, 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
DE LIGASE).
GN ARG OR PA3525.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO1:
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brickman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wedman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Labig G., Yuan Y.,
RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.B.W., Lory S., Olson M.V.;
RT "complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE -> AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC PATHWAY.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
DR EMBL: AE004773; AAC06913.1; -.
DR InterPro: IPR001518; Argininosuc_synth.
DR Pfam: PF00764; Argininosuc_synth.1.
DR ProDom: PD003544; Argininosuc_synth.1.
DR PROSITE: PS00564; ARGININOSUCCIN_SYN.1; 1.
DR PROSITE: PS00565; ARGININOSUCCIN_SYN.2; 1.
KW ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;
KW Urea cycle.
SQ SEQUENCE 405 AA; 45297 MW; 1C3DB39EB18689E5 CRC64;

```

```

alignment_scores:
Quality: 927.50 Length: 404
Ratio: 3.011 Gaps: 6
Percent Similarity: 76.238 Percent Identity: 47.030
alignment_block:
US-09-775-693-1 x Q9HY84 ..

```

Align seg 1/1 to: Q9H184 from: 1 to: 405

```

1  ANGTCCAGCAAGGCTCCGTGGTTCAGGCTACAGTGGCGGCTGGACAC 50
1  ::::::::::::::::::::::::::::::::::::::::::::::::::::
1  MetAlaAspValLysValValLeuAlaTyrSerGlyLysAsp 17
51  CTCGTGCATCCTCGTGTGAGCA...CAAGCTATGACGTGATG 97
51  ::::::::::::::::::::::::::::::::::::::::::::::::::::
17  rSerValIleLeuLysTrpLeuLysPheTyrAsnCysGluValT 34
98  CCAATCTGGCCACATTTGGCCAGAGACTTCGAGAGCCAGAG 147
98  ::::::::::::::::::::::::::::::::::::::::::::::::::::
34  hrPheThrAlaAspLeuGlyGlnGlyGluValAluProAlaArg 50
148  AAGGCACTGAGCTTGGGGCCAAAGAGTTCATTGAGAGTCCAG 197
148  ::::::::::::::::::::::::::::::::::::::::::::::::::::
51  LysAlaArgAlaMetGlyValLysGluIleTyrIleAspSer 67
198  GGAGTTTGTGAGAGTTCATCTGGCGGCGCATCCAGTCCAGCC 247
198  ::::::::::::::::::::::::::::::::::::::::::::::::::::
67  uGluPheValArgAspPheValTyrProMetLpPheArgAlaAs 84
248  ATGAGAGCCGCTACCTCTGGGCACTCTCTGCAAGGCCCTCAG 297
248  ::::::::::::::::::::::::::::::::::::::::::::::::::::
84  YrGluGlyGluTyrLeuLeuGlyThrSerIleAlaArgProLeu 100
298  CGCAACACAGTGAATGCCCGAGGGGAGGGGCCAGATGATGTC 347
298  ::::::::::::::::::::::::::::::::::::::::::::::::::::
101  LysArgLeuIleGlyIleAlaAsnGluThrGlyAlaAspAla 117
348  CGGGCCAGCAAGAGGGAAGCATGAGTCCGTTGAGTCACTGCT 397
348  ::::::::::::::::::::::::::::::::::::::::::::::::::::
117  sGlyAlaThrGlyLysLysAsnAspGlnValArgPheGluLeu 134
398  ACTGACTGGCCCCAGATTAAGGTCAATGCTCCCTGAGAGTCC 447
398  ::::::::::::::::::::::::::::::::::::::::::::::::::::
134  YrAlaLeuLysProGlyValLysValIleAlaProTyrArgGlu 150
448  TTGTACAAACCGGTTCAAGGGCCGCAATGACTGATGAGTACG 497
448  ::::::::::::::::::::::::::::::::::::::::::::::::::::
151  LeuLeuSerArgGlyLys.....LeuMetAspTyrAlaGly 163
498  ACAGGCGATTCCTCCATCCG.....GTCACTCCCAAGACCC 541
498  ::::::::::::::::::::::::::::::::::::::::::::::::::::
163  ShisGlyIleProIleGluArgHisGlyLysLysSerProTyr 180
542  TGGATGAGAACCTCAGATCAGATCAGAGGCTGAGATCCTGAG 591
542  ::::::::::::::::::::::::::::::::::::::::::::::::::::
180  eTAspAlaAsnLeuLeuHisIleSerTyrGluGlyValLeuGlu 196
592  CCCAAGAACCAAGCCCTCCAGGCTCTACAGAAACCCAGGACC 641
592  ::::::::::::::::::::::::::::::::::::::::::::::::::::
197  ThrTrpGluHisGluGluAspMetTyrTrpThrAlaSerPro 213
642  CAAGCCGCCCAACACCCCTGACATTCAGAGTCAAGTCAAAA 691
642  ::::::::::::::::::::::::::::::::::::::::::::::::::::
213  uAsnAlaProAspThrProThrTyrIleGluLeuThrTyrArg 230
692  TCCCTGTGAGGTGACCAACGTCAGATGCGACCAACCAGAC 741
692  ::::::::::::::::::::::::::::::::::::::::::::::::::::
230  sPrlaValAlaIle.....AspGly...LysAspMetThrPro 241
742  TTGGAGCTTCATGATGACCAAGAGTCGGGCGCAAGACAGG 791
742  ::::::::::::::::::::::::::::::::::::::::::::::::::::
242  AlaGluValLeuThrGluLeuAsnArgValGlyGlyIleAsn 258
792  CCGTATGACATCGTGAGAACCGCTTCATTGGAATGAAGTCC 841
792  ::::::::::::::::::::::::::::::::::::::::::::::::::::
258  YrGluLeuAspIleValGluAsnArgTyrValGlyMetLys 275
842  TCTACGAGACCCAGCAGACCATCTTACCATGCTCATTTAG 891
842  ::::::::::::::::::::::::::::::::::::::::::::::::::::

```

```

275  YsTyrGluThrProGlyGlyThrIleLeuLysAlaHisArgAla 291
892  GAGCGCTTCACCATGAGCCGGAAGTGGCAAAATCAACAGCC 941
892  ::::::::::::::::::::::::::::::::::::::::::::::::::::
292  GluSerIleThrLeuAspArgGluValAlaHisLeuLysAsp 308
942  CTTGAATTTGCTGAGCTGTGTATACCGGTTTACGGCTTAC 991
942  ::::::::::::::::::::::::::::::::::::::::::::::::::::
308  LProLysTyrAlaSerLeuIleTyrThrGlyTyrTrpSerPro 325
992  GTGAATTTGTCCGCGACATCGCCAGTCCCGAGGAGCGTGA 1041
992  ::::::::::::::::::::::::::::::::::::::::::::::::::::
325  rGluMetLeuGlnGlnMetIleAspAlaSerGlnValAsnVal 341
1042  AAGTGCAGGTGTCCTCTCAAGGCGCAGGTGATCATCTCGCC 1091
1042  ::::::::::::::::::::::::::::::::::::::::::::::::::::
342  ValValaGluLeuLysLeuTyrLysGlnValValaValaGly 358
1092  GTCCCGACTGCTCTCTCAATGAGAGCTGGTGCATGACGTG 1141
1092  ::::::::::::::::::::::::::::::::::::::::::::::::::::
358  sSerAspSerLeuPheAspAlaAsnIleAlaThrPheGluAs 375
1142  GTGAT...TATAGCCCACTGATGCCACCGGCTCATCAATCA 1188
1142  ::::::::::::::::::::::::::::::::::::::::::::::::::::
375  YrGlyAlaTyrAsnGlnAlaAspAlaAlaGlyPheIleLys 391
1189  CTCAGCGTGAAG 1200
392  LeuArgMetArg 395
seq_name: sp_bacteria:Q9PHK7
seq_documentation_block:
ID  Q9PHK7  PRELIMINARY;  PRT;  406 AA.
AC  Q9PHK7:
DT  01-OCT-2000 (TReMBLrel. 15, Created)
DT  01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT  01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE  ARGININOSUCCLINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
DE  LIGASE).
GN  ARG OR CJO665.
OS  Campylobacter jejuni.
OC  Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC  Campylobacter.
OX  NCBI_TaxID=197;
RN  [1]
RC  SEQUENCE FROM N.A.
RX  STRAIN=NCIC 11168;
RX  MEDLINE=20150912; PubMed=10688204;
RA  Parhill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,
RA  Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA  Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA  Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA  Whitehead S., Barrall B.G.;
RT  "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT  reveals hypervariable sequences."
RL  Nature 403:665-668(2000).
CC  -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE -> AMP +
CC  PYROPHOSPHATE + L-ARGININOSUCCLINATE.
CC  -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC  PATHWAY.
CC  -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC  -1- SIMILARITY: TO THE ARGININOSUCCLINATE SYNTHASE FAMILY.
DR  EMBL: AL139075; CAB75297.1; -.
DR  InterPro: IPR001518; Arginosuc_synth.
DR  Pfam: PF00764; Arginosuc_synth; 1.
DR  ProDom: PD003544; Arginosuc_synth; 1.
DR  PROSITE: PS00564; ARGININOSUCCLIN SYN 1; 1.
DR  PROSITE: PS00565; ARGININOSUCCLIN SYN 2; 1.
KW  ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;
KW  Urea cycle.
SO  SEQUENCE 406 AA; 45578 MW; 8A1E137AF30EC77F CRC64;

```

alignment_scores: Quality: 892.00 Length: 397
 Ratio: 3.014 Gaps: 6
 Percent Similarity: 74.559 Percent Identity: 47.355

alignment_block:
 US-09-775-693-1 x Q9PHK7 ..

Align seg 1/1 to: Q9PHK7 from: 1 to: 406

```

19 GTGCTGTGGCCCTACAGATGCGCGCCCTGGACACTCGTCGATCCTGTGTG 68
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8 ValValLeuAlaItyrSerItyrGlyLeuAspThrSerIleLeuLeuItyr 24
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 GCCAAGAGGAGACTTGCAGAGAGCCAGAGAGAGAGGAGGAGGAGTGTGG 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 LylngLylngLylngLylngLylngLylngLylngLylngLylngLylng 57
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
166 GCCAAAAG.....GTGTCATGTAGAGATGTCAGGAGGAGTGTGTGA 209
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
58 LlylsgLylngLylngLylngLylngLylngLylngLylngLylngLylng 74
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 GGAGTTCATGTGGCGCGCCATCCAGTCCAGCCAGCCAGTGTGTGAGACCGCT 259
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
74 SaspryIValPheProMetPheArgAlaIleItyrGlyLylngLylng 91
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 ACCTCTGGGACACTCTCTTGGCCAGGCCCTGTGATGCCCGCAAGAGTGT 309
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
91 LylLeuLeuLylLylSerIleAlaItyrProLeuLylAlaLylThrGlnAla 107
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
310 GAATTCGCCAGCGGGAGGGGCGCAAGTATGTCTGCCACGGCGCCACAGG 359
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
108 GlnIleAlaLeuGlnThrIleAlaIleAlaIleAlaIleAlaIleAlaIle 124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
360 AAGGAGGAGAGATGATCGGCTGTTGAGCTCAGTCTACACCTGAGGCGCC 409
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 LylsgLylngLylngLylngLylngLylngLylngLylngLylngLylng 141
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 CCCAATTAAGGTCTATGCTGCTCCCTGGAGAGTCCGTAATGTACACCGG 459
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 roAspLeuLylLylIleAlaItyrProItyrArgLylLylAspLeuAsnSerArg 157
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
460 TTCAAGGCGCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 509
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
158 GluLys.....LeuLeuAlaItyrAlaGlnLylShIsgLylIleAs 170
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
510 CATCCCGGTCACTCCC...AAGAACCCGTGAGAGATGATGATGATGATGAT 556
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 PileSerLylLylLylLylLylSerProLylSerMetAlaIleAlaIle 187
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
557 TGCACATAGTACGAGGCTGATGATGATGATGATGATGATGATGATGATGAT 606
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
187 euHisIleSerItyrGlyLeuValLeuGlnLylAspLeuValAlaIleAla 203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
607 COTCCAGGTCTCTACACGAGAGCCAGGAGCCAGGAGGAGGAGGAGGAGGAG 656
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 GluLylAspMetItyrArgItyrSerLylSerProLylAspAlaIleAla 220
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
657 CCCGACATTTCTCGAGATCGAGTTCAAAAAAGGGGCTCCCTGTGAAGGTGA 706
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
220 uSerGlnIleIleIleLylLeuAspPheGlnLylLylAspLeuValAlaIle 237
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
707 CCAAGTCAGAGATGAGCACCACCCAGACCAAGCTCTTGAGGAGCTTCATG 756
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 snGlyLylLys.....LeuSerProAlaGlyLeuLeuThr 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
757 TACCTGACGAGATGCGGCGCAAGCATGGCTGGCGCTATTCATCATCGT 806
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

249 LylLeuAlaIleGlnLylGlyLylShIsgLylIleGlyItyrArgLeuAspIleVala 265
807 GGAGAACCGCTTCATGTGAGATGAGTCCAGGAGTATTCACGAGACCCAG 856
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 LglLysnArgItyrValIglMetLylSerArgItyrGlyItyrGlnItyrProG 282
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
857 CAGGACCATCTCTTACCATGCTCATTTAGACATGAGGAGGAGGAGGAGGAG 906
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 LylGlyThrIleLeuLeuLylAlaIleHisArgAlaIleGlnItyrIleThrLeu 298
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
907 GACCGGAGAGTCCGCAAAATGCAAAACAGGCGCTGGAATTTGGCGA 956
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 AspArgGlnAlaIleAlaIleHisLeuLysAspIleLeuMetProLylItyrAlaIle 315
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
957 GCTGTGTATACCGGTTTACGCGCTAGCCCTGAGTGTGATTTGTCGCC 1006
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 LylLeuItyrAsnGlyItyrItyrPheSerProGlnArgMetIleGlnAla 332
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1007 ACTGCATGCCAAGTCCAGAGAGGAGGAGTGTGAGGAGGAGGAGGAGGAGTCC 1056
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 lAlaLeuIleAspIleSerIleIleHisAlaIleAsnGlyItyrValLylngLyl 348
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1057 GTCTCAAGGCGCAGGTGTATCATCTCGCGCGGAGTCCACATG...TC 1103
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 LeuItyrLylsgLylAsnValMetValIleGlyItyrGlnItyrAlaIleAsnAsp 365
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1104 TCTCTACATGAGAGCTGTGAGCATGAACTGAGGAGGAGGAGGAGGAGGAG 1153
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 LylLeuPheAsnAlaIleAlaItyrCysThrPheGlnLylngLylngLylngLylng 382
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1154 CAACGTATGCCAGCGGTTGATCAACATCAATTCATTCCTCCAG 1194
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 LylLysAspAlaIleAlaGlyPheIleLylLeuAsnAlaIleLeuArg 395
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: sp_bacteria:09K820

seq_documentation_block:

ID	09K820	PRELIMINARY:	PRT:	409 AA.
AC	09K820:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE LIGASE).			
GN	ARGG OR BH3187.			
OS	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=86665;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C-125 / JCM 9153;			
RX	MEDLINE=20512582; PubMed=11058132;			
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,			
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kohara S.,			
RA	Horikoshi K.;			
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus			
RT	halodurans and genomic sequence comparison with Bacillus subtilis."			
RT	Nucleic Acids Res. 28:4317-4331(2000).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +			
CC	PYROPHOSPHATE + L-ARGININOSUCCINATE.			
CC	-1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC			
CC	PATHWAY.			
CC	-1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).			
CC	-1- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.			
DR	EMBL: AP001518; BAB06906.1; -			
DR	InterPro: IPR001518; Arginosuc_synth.			
DR	Pfam: PF00764; Arginosuc_synth. 1.			
DR	ProDom: PD003544; Arginosuc_synth. 1.			
DR	PROSITE: PS00564; ARGININOSUCCIN_SYN_1. 1.			
DR	PROSITE: PS00565; ARGININOSUCCIN_SYN_2. 1.			
KW	ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;			
KW	urea cycle.			

```

seq_name: sp_bacteria:09K423
seq_documentation_block:
ID      09K423      PRELIMINARY;      PRT;      404 AA.
AC      09K423;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
GN      LIGASE).
GN      ARGG.
OS      Moritella sp. 2674.
OC      Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC      Moritella.
OX      NCBI_TaxId=111291;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=2674;
RX      MEDLINE=20158877; PubMed=10692366;
RX      Xu Y., Liang Z., Legrain C., Ruger H.J., Glandsdorff N.;
RT      "Evolution of Arginine biosynthesis in the bacterial domain: Novel
RT      gene-enzyme relationships from psychrophilic moritella strains
RT      (vibrionaceae) and evolutionary significance of n-alpha-acetyl
RT      ornithinase.";
RT      J. Bacteriol. 182:1609-1615(2000).
RT      RL
CC      -i- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC      PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC      -i- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC      PATHWAY.
CC      -i- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC      -i- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC      EMBL: AJ352020; CAB95017.1; -
DR      InterPro: IPR001518; Arginosuc_synth.

```

DR Pfam: PF00764; Arginosuc_synth. 1.
 DR Prodom: PD003544; Arginosuc_synth. 1.
 DR PROSITE: PS00564; ARGININOSUCIN_SYN_1.
 DR PROSITE: PS00565; ARGININOSUCIN_SYN_2; 1.
 DR AMP-binding: Arginine biosynthesis; Ligase; Urea cycle.
 KW SEQUENCE 404 AA; 44291 MW; EBB8AFB4E87F20 CRC64;

alignment_scores:

Quality: 818.00 Length: 401
 Ratio: 2.801 Gaps: 9
 Percent Similarity: 72.818 Percent Identity: 44.638

alignment_block:

US-09-775-693-1 x Q9K423 ..

Align seg 1/1 to: Q9K423 from: 1 to: 404

```

19 GTGCTTCTGCTACAGTGGCGCTGACACCTGTCGTCGCTGCTG 68
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 pLeuGlnGlu...AsnTyrAspAsnGlySerGluLeuAlaAla 42
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
69 GCTGAAGAGCAAGCTATGAC.....GTCATGCTCTATCGGCCA 109
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 pLeuGlnGlu...AsnTyrAspAsnGlySerGluLeuAlaAla 42
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
110 ACATGGGCGAG...AAGAACTTCGAGGAAGCCAGGAAGAGCAG 156
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
42 sPValGlyGlnGlyAlaGlnGlnGlnGlnGlnGlnGlnGln 58
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
157 AAGCTTGGGCGCAAAAGGTTCATTCAGAGATGTCAGACGAGCT 206
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 AlaserGlyAlaserGlyCysTyrValAlaAlaAlaAlaAla 75
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 GGAGAGCTCATCTGCGCGGCGCATCCAGTCACGACCTGATGAG 256
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 IGlusAsnTyrLeuTyrProthreuleuysThrGlyAlaVal 92
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
257 GCTACTCTCTGGGCGACCTCTCTGCGAGCGCTGATCGCCGCA 306
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 hTyrLeuLeuGlnGlyThrSerMetAlaArgProIleAlaAla 108
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 GTGGAATTCGCCGCGGAGGCGGAGGCGGAGGCGGAGGCGG 356
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 ValGluIleAlaArgLysValGlyAlaAlaAlaAlaAlaAla 125
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 AGGAAGGGAAGCATCAGTCCGCTTGGAGCTGCTGCTAC... 403
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 rGlyLysGlnLysAsnAspGlnIleArgPheGlu...SerCys 141
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 TGGCCCCCAGATTAAGTCAATGCTCCCTGGAGGATGCCGAT 453
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 eAlaIleProIleuThrValIleAlaProIleArgIleTyrP 155
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
454 AACCGGTTCAGAGCGCGCATGACCTGATGAGTGAAGCAAC 503
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 ....LeuThrSerArgLysLeuSerLeuGlnGlnGlnGln 170
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
504 GATTCCTCCATCCGCTCCTCCCAAGAACCCGCGGAGCATG 553
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 pIleProIleAlaAlaSerAlaThrLysIleTyrSerArgAsp 187
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
554 TCATGACATCAGCTACGAGCTGGAATCTGGAGAACCCCAAG 603
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 LatrPheIleSerHisGlnGlnGlnGlnGlnGlnGlnGln 203
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
604 GCGCTCCAGCTCTACAGCAAGAGCCAGGAGCCAGCAAGCC 653
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 ProSerLysGlnValThrMetThrValAspProIleAspAla 220
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
654 CACCCCGACATTCGAGATCGATTCAAAAAGGGGCTCCGTGA 703
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 nGluProGluPheLeuThrIleSerValValLysGly... 236

```

```

704 TGACCAACGTCACAGATGGACACCCAGACCTCTTGAGACTCTTC 753
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 lAlaAlaGlnGlnLysLysLeuSerProIleArgPhe...Leu 248
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
754 ATGATCTGACGAAGATCGCGGCGGAGCATGCGTGGCGCTGAT 803
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
249 MetTyrLeuAsnGlnLysAlaAlaAlaAlaAlaAlaAlaAla 265
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
804 CGTGAAGAACCGCTTCATTCGATGAGTCCCGGATGATCTAC 853
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 eValGlnAsnArgLeuValGlyMetLysSerArgGlyCysTyr 282
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
854 CAGCAGGCGACCATCTTACCATGCTCATTTAGACATCGAGCC 903
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 rGlyGlyThrValMetValGlnAlaAlaAlaAlaAlaAla 298
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
904 ATGACCGGGAAGTGGCGCAAAATCAACAGAGCTGGCTGAAT 953
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 LeuAspLysThrThrArgLysTyrLysGlnThrValAlaAla 315
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
954 TGAGCTGCTGATACCGCTTACGGCTGAGCTGATGATTTGTC 1003
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 rHisLeuValTyrAspGlyArgTyrPheThrProLeuGlyAla 332
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1004 GGCACCTGCATCCGCGCAAGTCCAGAGCGAGTGAAGGAAG 1053
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 eAlaAlaAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 348
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1054 TCCGCTCTCAAGGCGGAGGTGATACATCCGCGGAGGAGTCC 1103
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 LysMetTyrLysGlnSerValGlnAlaAlaAlaAlaAlaAla 365
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1104 TCCTACATGAGAGAGCTGCTGAGCATGAACGTCAGGCTGAT 1153
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
365 rLeuTyrSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 382
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1154 CAACGTATGCCAGCGGCTTCATCAATCAATTCCTC...AG 1197
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
382 InsertAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 398
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1198 AAG 1200
   |||
399 Lys 399

```

seq_name: sp.plant:Q9SZX3

seq_documentation_block:

ID Q9SZX3 PRELIMINARY; PRT: 498 AA.

AC Q9SZX3:

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE ARGININOSUCINATE SYNTHASE-LIKE PROTEIN (EC 6.3.4.5).

GN F617.40 OR AT4G24830.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RA Bevan M., Terry N., Ardiles W., Busschaert C., Dasseville R.,

RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,

RA Villarroel R., Gielens J., Van Montagu M., Bancroft I., Hohleisel J.,

RA Mewes H.W., Mayer K.F.X., Schueller C.,

RA Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.

RP (12)

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RA Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.

RL (13)

RP SEQUENCE FROM N.A.


```

663  CATTTCGAGATTCGAGTCACAAAAAGGGGCCCTGTCAGATGACCAAG 712
        : ||||| : : : : : ||||| : : : : :
315  wtyrllleglyllylleglusergylleuprovalAlAlaLeuAsnGly 331
        : ||||| : : : : : ||||| : : : : :
713  TCAGAGATGCGCACCCACCAACGCTCTTGGAGCTCTTCATGTCCTG 762
332  ..... LysAlaLeuSerProAlaThrLeuAlaGluLeu 343
        : : : : : ||||| : : : : :
763  AACGAATCGCGGCAACGATGGCGGCGCTTTCATCGNGAGAA 812
        : : : : : ||||| : : : : : ||||| : : : : :
344  AsnThrllleglyllyshisgyllylledlyrGlyLeuSmetValGluAs 360
813  CGGCTCATTTGATGATGATGCCGAGTCTTACAGACCCACGACGCA 862
360  natgLeuValGlyMetLysSerArgGlyValTylGluThrProGlyGly 377
863  CCATCTTTACCATGCTCATTTAGACATCGAGGCCCTTCACATGACGG 912
377  hrllleuPheAlaIaValGlnGlnIleuGlnSerLeuThrLeuAspArg 393
913  GAGTGCAGAAATCAACAGAGCGCTGGCTTGAATTTGCTGAGCGGT 962
394  GluSerllleGlnValLysAspThrLeuAlaLeuGlyTyrAlaGluMet 410
963  GTATACCGCTTACGCGCTAGCCCTGAGTGTGATTTTCCGCCACATGCA 1012
410  lTyrAlaGlyArgTyrPheAspProLeuArgGluSerMetAspAlaPheM 427
1013  TCGCCAAATCCAGAGCGAGGAAAGGAAAGTGCAGATGCGGCTCTC 1062
427  etGluTylserlIleThrGluThrThrGlySerValThrLeuLysLeuTyr 443
1063  AAGGCGCAGGTGATACATCCTCGGCCGGAGTCCCACTGCTCTTACAA 1112
444  LysGlySerValSerValThrGlyArgGlnSerProAsnSerLeuTyrAr 460
1113  TGAGAGCTGCTGAGCATGACATGACGGGTGATATAGCCCACTGATG 1162
460  gGlnAspIleSerSerPheGlnGlySerGluTleTyrAsnGlnAlaAspA 477
1163  CCACGGGTCATCAACATCATTCATCCTCAGGCTGAG 1200
477  lAlaAlaGlyPheIleArgLeuTyrGlyLeuProMetLys 489

seq_name: sp_bacteria:09k4Y8

seq_documentation_block:
ID 09k4Y8; PRELIMINARY; PRT: 404 AA.
AC 09k4Y8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE-ASPARTATE
DE LIGASE).
GN ARGG.
OS Moritella sp. 2693.
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC Moritella.
OC Moritella.
OX NCBI_TaxID=111292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2693.
RA MEDLINE=20158877; PubMed=10692366;
RX Xu Y., Liang Z., Legrain C., Rieger H.J., Glandsdorff N.;
RT "Evolution of Arginine biosynthesis in the bacterial domain: Novel
RT gene-enzyme relationships from psychrophilic moritella strains
RT (vibrionaceae) and evolutionary significance of n-alpha-acetyl
RT ornithinase ";
RL J. Bacteriol. 182:1609-1615(2000).
CC -I- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARATE = AMP +
CC -I- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC

```

CC PATHWAY.
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
 DR EMBL: AJ252021; CAB95023.1; -
 DR InterPro: IPR001518; Arginosuc_synth.
 DR Pfam: PF00764; Arginosuc_synth.1.
 DR ProDom: PD003544; Arginosuc_synth.1.
 DR PROSITE: PS00564; ARGININOSUCCIN_SYN_1; 1.
 DR PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1.
 KW ATP-binding; Arginine biosynthesis; Ligase; Urea cycle.
 SQ SEQUENCE 404 AA; 44432 MW; 9EIC4D037AF5923F CRC64;

alignment_scores:
 Quality: 776.00 Length: 403
 Ratio: 2.713 Gaps: 10
 Percent Similarity: 70.968 Percent Identity: 43.672

alignment_block:
 US-09-775-693-1 x Q9K4Y8 ..

Align seg 1/1 to: Q9K4Y8 from: 1 to: 404

```

19 GTGGTTCTGGCCCTACAGTGGCGGCGTGGACCTGTCGATCTCTGTG 68
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
8 ValValValValValValValValValValValValValValValVal 24
69 GCTGAAGGACAAAGCCTATGAC.....GTCATTCCTTCCTGCG 106
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
24 pLeuGluGlu...AsnTyrAspAsnCysGluLeuValLeuValLeu 40
107 CCAACATTTGGCCAG...AAGGAAGACTTCGAGGAAGCCAGGAAGCA 153
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
40 LaAspValAlGluGluGluGluGluGluGluGluGluGluGluGlu 56
154 CTGAAGCTTGGGGCCAAAGGTTTCATTGAGATGTCACAGGAGGATT 203
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 LeuAlaSerGlyAlaSerGlyCysTyrValValAlaSerLeuLys 73
204 TGTGAGAGTTTCATTCGCGCGGCGGCGGCGGCGGCGGCGGCGGAG 253
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
73 uValGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 90
254 ACCGCTACCTCTGGGACCTCTCTGTCGAGGCGGCGGCGGCGGCGG 303
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90 LysThrTyrLeuLeuGlyThrSerMetAlaArgGlnSerIleAla 106
304 CAAGTGAATTCGCCCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 353
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 GlnValGluIleAlaArgLysValAlaAlaAspAlaLeuGluGlu 123
354 CACAGGAAGGGAAGCATGAGTCCGTTTGTGAGCTCAGCTGCTAC...T 400
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 sThrGlyLysGlyLysAspGlnIleArgPheGlu...SerCysPheAla 139
401 CACGCGCGCGCGGAGATTAAGTATTCCTCTGAGGAGATGCTGAATTC 450
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
139 ILeuAlaProGluLeuThrValIleAlaIleProIleArgIleTyrP 154
451 TACAACGGTTCAAGGGCGGCGGAGTACGATGAGTACGCAAGCAACA 500
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 .....LeuThrSerArgGluSerLeuLeuGluIleAlaGluAla 168
501 CGGAGATTCCTCCATCCCGTCACTCCCAAGAACCGGTGAGATGATGAG 550
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 gaSPiLeProThrAlaAlaSerGlyThrIleTyrSerArgAspAla 185
551 ACCTCATGCATCATCATCATCATCATCATCATCATCATCATCATCAT 600
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
185 smaIaTrrPhtSileSerHisGluGlyGluLeuGluLysProIleP 201
601 CAAGCGCTCCAGTCTCTACAGGAAGGAGGAGGAGGAGGAGGAGGAG 650
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

202 GlnProSerLysGluValIleThrMetThrValAspProIleAspAla 218
651 CAACACCCCTGACATTCCTGAGATGAGTTCACAAAAGGGCTCCCTGCA 700
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 oAsnGluProGluPheLeuThrIleSerValValIleGly.....G 232
701 AGGTACCAACAGCTCAAGAGTGCACACCACACAGACCTCTGTGAGCTC 750
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
232 IuIleThrAlaValAsnGluGluGluMetTyr.....LeuIleLea 246
751 TTCATGTACCTGACAGAGTGCAGGCGGCGGCAAGCATGCGGCGCTAT 800
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 TyrThrTyrLeuAsnGluLysAlaAlaAlaHisGlyValGlyArgVal 263
801 CATGCTGAGAACCGCTTCATTGATGAATGAAGTCCCGAGGTTCATCAG 850
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 pIleValGluAsnArgLeuValGlyMetLysSerArgGlyCysTyrIle 280
851 CCCCAGCAGCAGCATCTTACCATGCTCATTTAGACATGAGCGCTTC 900
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
280 hrProGlyGlyThrValMetValGluAlaLeuArgGlyIleGluGlu 296
901 ACATGACCGGGAAGTGGCCAAATCAACAAGCGCTGGCTTGAATT 950
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
297 ValLeuAspLysIleThrArgLysTrrPlyHisThrValAlaAlaGlu 313
951 TGCTGACCTGGTGTATACCGGTTACGCGCTACCGCTGATGATGAA 1000
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 eSerHisLeuValTyrLysPheArgTrpPheThrProLeuGlyAlaSer 330
1001 TCCGCGCATCGATCGCAATCCAGAGAGGAGTGAAGGAAGTGCAG 1050
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
330 euLeuAlaAlaAlaGlyThrLeuAlaGluIleMetAsnGlyValIle 346
1051 GTGTCCGCTCTCAAGGCGCAGGTGTACATCTCGCGCGGAGGCCACT 1100
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 ValLysMetLysGlySerValGlnAlaValGlnLysGlnSerPro 363
1101 GTCTCTACATGAGAGAGTGTGAGCATGAAGCTGCAGGCTGATT 1149
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
363 nSerLeuTyrSerGluGluPheAlaThrPheGlyAspAsnValTyrA 380
1150 ..GAGCCACTGATGCCACCGGTTTCATCAACATCAATTCCTTC 1191
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
380 spAspIleSerHisAlaGluGlyPheIleArgLeuTyrSerLeuSer 396
1192 AGGCTGAAG 1200
397 ArgIleLys 399

```

seq_name: sp_bacteria:Q99VC7

seq_documentation_block:

ID Q99VC7 PRELIMINARY; PRT; 401 AA.

AC Q99VC7;

DT 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE ARGININOSUCCINATE SYNTHASE.

GN ARG OR SA0822.

OS Staphylococcus aureus subsp. aureus N315.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Staphylococcus.

OX NCBI_TaxID=158879;

RN [1]

RP SEQUENCE FROM N.A.

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,

RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Uji Y.,

RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,

RA Hirakawa H., Kohara S., Goto S., Yabuzaki J., Kanehisa M.,

RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,

RA Ogasawara N., Hayashi H., Hiramatsu K.;


```

alignment_scores:
  Quality: 751.50      Length: 394
  Ratio: 2.674         Gaps: 6
  Percent Similarity: 71.320      Percent Identity: 41.878
alignment_block:
  US-09-775-693-1 x Q9AKNT8 ..

```

[illegible]

```

seq_documentation_block:
ID      09CC10;          PRELIMINARY;          PRN;          399 AA.
DT      01-JUN-2001 (TREMBLrel. 17, Created)
DT      01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      ARGINOSUCCINATE SYNTHASE.
GN      ARGG OR ML1412.
OS      Mycobacterium leprae.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OX      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RX      NCBI_TaxID=1769;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=TN.
RX      MEDLINE=21128732; PubMed=11234002.
RA      Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA      Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA      Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
RA      Holroyd S., Horsley T., Jagers K., Lacroix C., Maclean J., Moule S.,
RA      Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.W.,
RA      Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA      Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,

```

RA Barrell B.G.: "Massive gene decay in the leprosy bacillus."

alignment_scores:	length:
Quality:	685.50
Ratio:	2.422
Percent Similarity:	68.357
	Percent Identity:
	37.440

Align seg 1/1 to: Q9CC10 from: 1 to: 399

[illegible]

```

648 CCCACACCCCTGCATCTCTGCAGATCGAGTCCAAAAAAGGCGTCCG 637
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212 ntrpAstrhrproAserbglu1llevalglrphglnslglsvalrprov 229
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
698 tgaAGGACGACCAAGTCGAAGATGGCGACCCACCCACAGACCTCTTGAG 747
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
229 alserlle.....AspGlySer..ProValSerMetIleuGly 241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
748 CTTCTCATGTACTGACGAAGAGTCGCGGACCAAGAGATGCGTGGCCGAT 797
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 lalleeluala..leuAsnArgArgAlaGlyAlaGlnGlnAlaGlyArg 257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
798 TGACATCTGTGGAGAACCGCTTCATGTGGAAATGAAGTCGCGAGGTATCGAG 847
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
257 uasPValValGlnAspArgleuValGlyIleuGlySerArgGlyIleuGly 274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
848 AGACCCGACGAGGACACATCTTTACACATCGCTATTTAGACATCGAGGCC 897
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
274 lualrProgluAlaMetValIleuIleThrAlaHisAlaGlnIleuGlnHis 290
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
898 TTGCACATGGACCGGAGATGGCCAAATCAACAAAGGCGCTGGGCTTGAA 947
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
291 valrhrleuGluIaTgIleuIleuArgPheIleuGlyArgIleThrAspArg 997
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
948 ATTTCCTAGCTGGTGTATACGCGTGTTCAGGCTACCCCTGAGTGTGAT 997
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
307 gtrpAlaGluIleuValItyrAspItyleuItrPtyrSerProIleuGlyThr 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
998 TTTCGCGCATCGATCGCCAAATGCCAGGACGAGCATGTCGAAGGAAGTG 1047
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 lalenuIserPheValAlaAlaThrGlnGlnIleValIthrGlyVal 340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1048 CAGGTGCTGCTCTCAAGGCGCAGGTGTATCTGCGGCGGAGATCCCC 1097
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
341 ArgMetValIleuIleGlyOlnHisIleAlaValAlaAsnGlyAlaArgSerAl 357
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1099 ACTGCTCTCTCAATGATGAGACTGTGTACATGAACATGCGACGGTGAT 1147
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
357 agIuIserIleuItyrAspPheAsnIleuAlaIthrTyrAspGluGlyAspThr 374
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1148 ATGACGCAATGATGCCACCGGGTTCATCAATCAATCTCCCTCAGGCTG 385
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 heAspGlnSerAlaAlaAlaArgGlyPheValItyVal..... 385
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1198 AAGGATATCATCGCTGTCCAGACCAAGTCACCTGGCCAAA 1236
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
386 .....TyrGlyIleuProSerItyrIleuAlaAlaAlaArg 395
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: sp-archea:Q9UX31

seq_documentation_block:
ID Q9UX31 PRELIMINARY; PRT; 406 AA.
DC Q9UX31;
DE 01-MAY-2000 (TREMBLrel_13, Created)
DT 01-MAY-2000 (TREMBLrel_13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel_11, Last annotation update)
DE ARGININOSUCINATE SYNTHASE (EC 6.3.4.5).
DI
DR ARGININOSUCINATE SYNTHASE
GN Sulfolobus solfataricus; Sulfolobaceae; Sulfolobus.
OS Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID-2287;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1617 / P2;
RC Charlebois R.L., Singh R.K., Chan-Weiher C.C.-Y., Allard G., Chow C.,
RA Gasterlinder F., Curtis B., Duguet M., Eruso G., Fanny D.,
RA Gasterlinder F., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,
RA Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She O.,
RA St Jean A., Van Der Oost J., Young F., Zivanovic Y., Doolittle W.F.,
RA Ragan M.A., Sensen C.W.;
RT "gene content and organization of a 281-kbp contig from the genome of
the extremely thermophilic archaeon, Sulfolobus solfataricus P2."

```

Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y18930; CAB57663.1; -
 DR InterPro: IPR001518; Argininosuc synth.
 DR Pfam: PF00764; Argininosuc synth_1.
 DR ProDom: PD003544; Argininosuc synth_1.
 DR PROSITE: PS00564; ARGININOSUCIN_SYN_1; UNKNOWN_1.
 KW LIGASE.
 SEQUENCE 406 AA; 46285 MW; 3956BABC9FC04C12 CRC64;

alignment_scores:

Quality: 656.50 Length: 394
 Ratio: 2.441 Gaps: 9
 Percent Similarity: 68.274 Percent Identity: 39.594

alignment_block:

US-09-775-693-1 x Q9UX31 ..

Align seg 1/1 to: Q9UX31 from: 1 to: 406

```

19 GTGTTCTGCGCTACAGTGGCGGCTGGAGACCTCGTCACTCGTGTG 68
   ::::::::::::::::::::::::::::::::::::::::::::
18 IIEVALLLEUALATYSEKGLYLLEUASPRHTRHVALSERILEAGTR 34
   ::::::::::::::::::::::::::::::::::::::::::::
69 GCTGAAGGAACA...GGCTATGACGTTCATTGCTTGGCCACATTTG 115
   ::::::::::::::::::::::::::::::::::::::::::::
34 PLEUYSGLHTRHRELHLSALAGLUILETHRVALTHRVASLSPVIG 51
   ::::::::::::::::::::::::::::::::::::::::::::
116 GCCAAGAGAGACTTGGAGAGCCAGAGAGAGAGAGAGAGAGAGAG 165
   ::::::::::::::::::::::::::::::::::::::::::::
51 LYLINLSPSPSPRHELYSLYILEGLGLATGATGATGATGATGATG 111
   ::::::::::::::::::::::::::::::::::::::::::::
166 GCCAAAAGGTTCATTGAGATGTCAGACAGAGAGAGAGAGAGAG 215
   ::::::::::::::::::::::::::::::::::::::::::::
68 ALASERLYSHSTRHTRHLEASPLALVALARGLINPHEALASNSNTY 84
   ::::::::::::::::::::::::::::::::::::::::::::
216 CATCTGCGCGGCGCATCCAGGCGCATCTATGAGAGAGAGAGAG 265
   ::::::::::::::::::::::::::::::::::::::::::::
84 RLLEALATYRALATLLELYSLLEUASNGLYLEUTRYGLNGLYVALTY 101
   ::::::::::::::::::::::::::::::::::::::::::::
266 TGGGCGACCTCTCTTGGCGGCGCTGATGCGCGGCGGCGGCGG 315
   ::::::::::::::::::::::::::::::::::::::::::::
101 ELSETHRLALEUALLARGPROLEULLEALILELYSVALALGLUVAL 117
   ::::::::::::::::::::::::::::::::::::::::::::
316 GCCAGCGGAGAGGCGGCGGAGTGTGTCCAGCGGCGGCGGAGAG 365
   ::::::::::::::::::::::::::::::::::::::::::::
118 ALALYSLSGLNGLYALAGLUALLALALANLISGLYSERHTRHSE 134
   ::::::::::::::::::::::::::::::::::::::::::::
366 GAAGATCAGTCCGCTTGGAGTCACTGCTACCTGCTGCTGCTGCT 415
   ::::::::::::::::::::::::::::::::::::::::::::
134 YASNSPGLINVALATGPHASPLLEUALLVALLYSALALEUUTYR 456
   ::::::::::::::::::::::::::::::::::::::::::::
416 TAAAGTCACTGCTGCC...TGGAGGATGCGCTGATTCCTACAC 456
   ::::::::::::::::::::::::::::::::::::::::::::
151 ALLYSLEILELLEALPROALARGILETRHSPNSMETHR..... 163
   ::::::::::::::::::::::::::::::::::::::::::::
457 CGGTTCAAGGCGCGCATGACCTGATGATGAGTACCAAGAGAGAG 506
   ::::::::::::::::::::::::::::::::::::::::::::
164 .....ARGGLASPLGLILELYSTYALALYGLILYGLIYL 176
   ::::::::::::::::::::::::::::::::::::::::::::
507 TCCGATCCCGGCTACTCCCAAGAACCGGTGAGCATGATGAGAAC 556
   ::::::::::::::::::::::::::::::::::::::::::::
176 EPRIOLLEYVAL...GLUSERSPLYSYRISERILEASPLINSLEUT 192
   ::::::::::::::::::::::::::::::::::::::::::::
557 TGCACATCAGTCAAGGCGGAGATCTGAGAACCCCAAGAACGAG 606
   ::::::::::::::::::::::::::::::::::::::::::::
192 RPELYATSERILEGLIYASPLILELLEASERASPROSERILEGLI 208
   ::::::::::::::::::::::::::::::::::::::::::::
607 CTTCCAGGCTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 656
   ::::::::::::::::::::::::::::::::::::::::::::
209 PROGLIASPALAHNEGILUTRPHLYS.....GLINLETYRASNLY 222

```

```

657 CCCTGACATCTTCGAGATCGAGTTCAAAAAAGGGTCCCTGTAAGTGA 706
   ::::::::::::::::::::::::::::::::::::::::::::
222 SLYSGLILEVALSERILEGLIUPHESERASNGLYVALPROTHRALVAL 239
   ::::::::::::::::::::::::::::::::::::::::::::
707 CCAACGTCAAGATGATGCGACCAACCCAGACCTCTCTGAGCTCTTCAT 250
   ::::::::::::::::::::::::::::::::::::::::::::
239 SNGLYGLIULYS.....METSGLLEUASNLYSLEUVALASP 250
   ::::::::::::::::::::::::::::::::::::::::::::
757 TACCTGAACGAGATCGCGGCGGAGCATGGCGGCGGCTATGATCGT 806
   ::::::::::::::::::::::::::::::::::::::::::::
251 PHELEUASNLEULYSRHEGLYSERHISGLYRHEGLYARVALIGLHIS 267
   ::::::::::::::::::::::::::::::::::::::::::::
807 GGAGAACCGCTTCATGGAATGAGTCCGAGATGATGAGAGAGAGAG 856
   ::::::::::::::::::::::::::::::::::::::::::::
267 EGIUASNARVALVALGILYRHELYSERARGGLUVALTYGLIUALPROA 284
   ::::::::::::::::::::::::::::::::::::::::::::
857 CAGGCGCATCTTTACATGCTCATTTAGACATCGAGGCTTCACCATG 906
   ::::::::::::::::::::::::::::::::::::::::::::
284 LALALEUCLYLEULIETRYALHISLLEASPLEUCLYSTRHILETYR 300
   ::::::::::::::::::::::::::::::::::::::::::::
907 GACCGGGAAGTGGCAAAATCAACAGGCTTGGGCTTGAATTTGCTGA 956
   ::::::::::::::::::::::::::::::::::::::::::::
301 THRPROMETGLULEUATGRHELYSARGNLSILEASPLINLEUTRSPER 317
   ::::::::::::::::::::::::::::::::::::::::::::
957 GCTGTGATACCGGCTTTACGCGCTGAGCGGCTGATGATGATTTGCG 317
   ::::::::::::::::::::::::::::::::::::::::::::
317 PLEUVALTYRGLNGLYLEUTRPHGLUPROLEUARGGLUTRHEULHIS 334
   ::::::::::::::::::::::::::::::::::::::::::::
1007 ACCTGATCGGCGGAGTCCGAGGAGAGAGAGAGAGAGAGAGAGAG 1056
   ::::::::::::::::::::::::::::::::::::::::::::
334 YSVALALASPLUMELASNLYSTRIPLESERGLYGLIUALYSVALIG 1056
   ::::::::::::::::::::::::::::::::::::::::::::
1057 GTCTCTAAGCGGCGGAGTACATCTCGCGGAGAGAGAGAGAGAGAG 1106
   ::::::::::::::::::::::::::::::::::::::::::::
351 VALSERANGLYSERPHALGLILEVALIYARGLISERGLUTRYSER 367
   ::::::::::::::::::::::::::::::::::::::::::::
1107 CTACATATGAGAGAGTGTGAGCATGAACGTCGAGGTCATGAGAG 1156
   ::::::::::::::::::::::::::::::::::::::::::::
367 OLYSERGLIULYSLEALASERTRYASN...LYSLYTRTYR...PROS 382
   ::::::::::::::::::::::::::::::::::::::::::::
1157 CTGAT.....GCCACCGGCTGATCATCAATC 1182
   ::::::::::::::::::::::::::::::::::::::::::::
382 ERASPLUMETALALARGGLYRHEILEGLIULI 392

```

seq_name: sp_human:043348

seq_documentation_block:

```

ID 043348 PRELIMINARY; PRT: 166 AA.
AC 043348:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE HYPOHETICAL PROTEIN RG007015.1 IN CHROMOSOME 7Q31 (FRAGMENT).
GN RG007015.1
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Stoneking T., Langston Y., Ahrens C.;
RU Submitted (Jan-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC003989; AAB96328.1; -
DR InterPro: IPR001518; Argininosuc synth.
DR Pfam: PF00764; Argininosuc synth_2.
DR ProDom: PD003544; Argininosuc synth_1.
KW Hypothetical protein.
FT NON TER
SEQUENCE 166 AA; 18967 MW; 26B6A3ADCBEFFAD CRC64;

```

alignment_scores:

Quality: 604.50

Length: 185

Ratio: 4.003 Gaps: 2
Percent Similarity: 81.622 Percent Identity: 67.568

alignment_block:
US-09-775-693-1 x 043348 ..

Align seg 1/1 to: 043348 from: 1 to: 166

```

682 AAAAAAGGGGTCCTGTGAGAGTGCACCAAGTCAGATGACACCA 731
1 LysLysGlyValProValLysValThr...LleLysMetAlaProPronH 16
732 CCAGACCTCTGTGAGAGTGCATACCTGACAGAACTCCGGGCAAC 781
16 sctlnhrSerLeuGluLeuPheValTyrLeuAsnGluValThrGlyLysH 33
782 ATGGCGTGGCGCGATTTGACATCGGTGAGACCCCTTCATGGAATGAG 831
33 lsglYmeGlyCysIleAspIleMetGluAsnCysPheIleArgIleLys 49
832 TCCGAGGTATCTACGAGACCCACAGACGACATCTTACATGCTCA 881
50 SerAlaGlyIleTyrLysThrProAlaGlyThrIleLeuGlnAlaAs 66
882 TTTAGACATCGAGGCTTCACCATGACCGGGAAGTCCGCAAAATCAAC 931
66 pLeuAspIleGluAspPheThrMetAspArgGluValHisLysIle... 81
932 AAGGCTGTGGCTTGAATTTGCTGAGCTGTGATACCGGTTACGGCT 981
81 ..... 81
982 AGCCCTGAGTGAATTTGCCGACATGCATGCCAAGTCCAGAGAGCG 1031
82 AsnProGluCysLysSerValHisHisCysIleThrLysSerGlnGlu 98
1032 AGTGAAGGGAAGTGCAGGTGTCCGCTCAAGGCGGCGAGTTCATATCC 1081
98 nmeGluGlyLysValGlnValProIlePheLysGlyGlnValHisLys 115
1082 TCGGCGGAGAGTCCCGACGTCTCTACATAGAGAGCTGTGAGCATG 1131
115 eusSerGlnGluSerLeuLeuSerLeuTyrLysGlnGluLeuValSerVal 131
1132 AACGTGAGGAGTATTTAGACCACTGATGCCAGCGGTTGATCAACAT 1181
132 AsnValGlnGlyAsnTyrGluProIleAspAlaThrGlyPheIleAsn 148
1182 CAATTCCTCAGGCTGAGAGATATATCATGCTCCAGACAGAGTCACTG 1231
148 eAsnTyrLeuArgLeuLysGluTyrHisTyrPheGlnSerLysValThr 165
1232 CCAAA 1236
165 alyys 166

```

seq_name: sp_archaea:Q9HMQ2

seq_documentation_block:
AC Q9HMQ2 PRELIMINARY; PRT: 396 AA.
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE ARGININOSUCCLINATE SYNTHETASE.
GN ARGG OR VNG2437G.
OS Archaea: Euryarchaeota, Halobacteriales; Halobacteriaceae;
OC Halobacterium.
NCBI_TaxID=64091;
RX [1]
RP SEQUENCE FROM N.A.
MEDLINE=20504483; PubMed=11016950;

```

RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angerline C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005124; AAG20519.1; -.
DR InterPro; IPR001518; Arginosuc-synth.
DR Pfam; PF00764; Arginosuc-synth; 1.
DR ProDom; PD003544; Arginosuc-synth; 1.
DR PROSITE; PS00564; ARGININOSUCCLIN_SYN_1; UNKNOWN_1.
DR PROSITE; PS00565; ARGININOSUCCLIN_SYN_2; 1.
KW Complete proteome.
SQ SEQUENCE 396 AA; 41994 MW; 8DA8C2DCAD94D092 CRC64;

```

alignment_scores:

Quality: 539.50 Length: 424
Ratio: 2.107 Gaps: 14
Percent Similarity: 60.377 Percent Identity: 33.255

alignment_block:
US-09-775-693-1 x Q9HMQ2 ..

Align seg 1/1 to: Q9HMQ2 from: 1 to: 396

```

13 GGCTCCGTGTTCTGCGCTACAGTGGCGGCTGGACACTCG...TGCAT 59
4 GlyThrValAlaLeuAlaPheSerGlyGlyLeuAspThrValGlyVal 20
60 CCTGTGCTGCTGAGAGACAGGCTAT...GACGTCATTCCTATCTG 106
20 LProLeuLysGlnGluTyrGlyTyrAspArgValIleGlyLysThr 37
107 CCACATTTGGCGCAAGGAA...GACTTCGAGAGACCCAGAGAAAGCA 153
37 AlaAspValGlyGlnProAspAlaGluPheAlaAlaAlaArgIleThrAla 53
154 CTGAAGCTTGGGGCCAAAAGTGTTCATTTAGAGATGCAGACGAGATT 203
54 AlaAlaLeuGlyValGluHis...HisValValAspAlaThrAlaGluPh 69
204 TGTGGAGAGATTGATCTGCGCGGCAATCCAGTCCAGCGCATGTGAGG 253
69 eAla...AlaLeuCysPheAspAlaValArgAlaAsnAlaThrTyrGln 85
254 ACCGCTACTCTGTGGCACCTCTGTGGCAGGCGCTGCATGCCCGCAAA 303
85 Ly...TyrProLeuGlyThrAlaLeuAlaArgProValIleAlaAspAla 100
304 CAAGTGGAAATCCGCCAGCGGAGGGGCCAAGTATGTGTCCACAGCGCC 353
101 IleLeuSerValAlaGluAlaGluGlyCysAlaAlaLeuAlaHisGlyCy 117
354 CACAGAAAGGGGAAAGATCAGTCCGGTTGAGTCAAGTCTGCTACTGAC 403
117 sThrGlyLysGlyAsnAspGlnLeuArgPheGlu..... 128
404 TGGCCCCCAGATAAAGTCAATGCTCCCTGGAGAGATCCGTAATTTCAC 453
129 .....AlaValTTPArgAlaSerAsp..... 135
454 AACCGTTCAAGGCGCGCATATGCTG..... 480
136 .....HisAspValCysAlaProValArgGluLeuG 146
481 .....ATGAGTACCAAGACACAGACAGCTTCCCA 511
146 yLeuThrArgGluTyrGluIleGlyTyrAlaAlaGluArgAspLeuPro 163

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:44:27 ; Search time 1806.06 Seconds
(without alignments) 11317.456 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239
Sequence: 1 atgtccagcaagagcttcgt.....gcaaggtcactgcgaataag 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb_ba: 2: gb_hcg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vi: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_om: 20: em_or: 21: em_ov: 22: em_pat: 23: em_ph: 24: em_pl: 25: em_ro: 26: em_sts: 27: em_sy: 28: em_un: 29: em_vi: 30: em_hgo_hum: 31: em_hgo_inv: 32: em_hgo_rod: 33: em_hgo_hum: 34: em_hgo_inv: 35: em_hgo_rod: 36: em_hgo_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	100.0	1560	9 HSASD	X01630 Human mRNA
2	1232.6	99.4	1975	9 AK027126	AK027126 Homo sapi
3	1231	99.4	1595	9 BC009243	BC009243 Homo sapi
4	1093	88.2	180838	2 AC027153	AC027153 Homo sapi
5	1091.4	88.1	165439	2 AC025175	AC025175 Homo sapi
6	1067.8	86.2	1398	4 BOVASS	M26198 Bovine ASS
7	1058	85.4	188356	2 AC009550	AC009550 Homo sapi
8	1058	85.4	190665	2 AP003057	AP003057 Homo sapi
9	1058	85.4	201989	2 AC073169	AC073169 Homo sapi
10	1057.8	85.4	155019	2 AL353717	AL353717 Homo sapi
11	1055.4	85.2	159687	2 AC087311	AC087311 Homo sapi
12	1055.4	85.2	204371	2 AC026357	AC026357 Homo sapi
13	1052.8	85.0	107314	9 AC005000	AC005000 Homo sapi
14	1052.2	84.9	1868	9 HUMAS3PS	AC008717 Homo sapi
15	1032	83.3	110716	2 AC008717	AC008717 Homo sapi
16	1031.8	83.3	83898	2 AC078791	AC078791 Homo sapi
17	1030.4	83.2	197429	2 AC023666	AC023666 Homo sapi
18	1030	83.1	58408	2 AL160400	AL160400 Homo sapi
19	1030	83.1	151833	2 AC016875	AC016875 Homo sapi
20	1027.8	83.0	1495	10 RATASSA	AL160400 Human DNA
21	1027.8	83.0	1495	10 RNASS	AC016875 Homo sapi
22	1016.6	82.1	1645	10 MUSASSB	M36708 Nucleotide
23	1016.6	82.1	1645	10 MUSASSB	X12459 Rat mRNA fo
24	1014	81.8	1877	9 HUMASALPS	M31690 Mouse argin
25	1011.2	81.6	149752	2 AC004616	BC002074 Mus muscu
26	1009.6	81.5	160658	2 AC073582	K01845 Human argin
27	988.2	79.8	113799	2 AC012502	AC004616 Homo sapi
28	988.2	79.8	173690	2 AC012502	AC073582 Homo sapi
29	986.2	79.6	176948	2 AL353705	AC012502 Homo sapi
30	986.2	79.6	194389	2 AC018693	AC011307 Homo sapi
31	961	77.4	347974	2 AC092833	AL354700 Homo sapi
32	959.4	77.4	347974	2 AC092833	AL353705 Homo sapi
33	570.8	46.1	132930	2 AC079046	AC092833 Homo sapi
34	567	45.8	81346	2 AC021496	AC079046 Homo sapi
35	551	44.5	808	9 HUMAS08	AC021496 Homo sapi
36	536.4	43.3	179658	2 AC024525	I00084 Human argin
37	427.4	34.5	10007	1 AE001816	AC024525 Homo sapi
38	419.4	33.8	81346	2 AC021496	AE001816 Rhesuslog
39	362.8	29.3	118103	9 AC003989	AC021496 Homo sapi
40	325.6	26.3	10040	1 AE001924	AC003989 Homo sapi
41	321	25.9	336	1 HUMAS05	AE001924 Delnoco
42	315	25.4	10138	1 AE005687	I00081 Human argin
43	310.4	25.1	360	6 AX192950	AE005687 Caulobact
44	305.2	24.6	405	9 AX192950	AX192950 Sequence
45	301.6	24.3	41255	8 SPBC428	D16950 Human HepG2

ALIGNMENTS

RESULT 1
HSASD 1560 bp mRNA PRI 12-SEP-1993
DEFINITION Human mRNA for argininosuccinate synthetase.
ACCESSION X01630.1 GI:28871
VERSION X01630.1 GI:28871
KEYWORDS synthetase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1560)
AUTHORS Bock,H.G., Su,T.S., O'Brien,W.E. and Beaudet,A.L.
TITLE Sequence for human argininosuccinate synthetase cDNA
JOURNAL Nucleic Acids Res. 11 (18), 6505-6512 (1983)
MEDLINE 84015388
FEATURES
location/Qualifiers
1..1560
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

misc_feature      40..48
                  /note="three tandem arginin codons"
CDS               76..1314
                  /note="argininosuccinate synthetase (aa 1-412)"
                  /codon_start=1
                  /protein_id="CAA25771.1"
                  /db_xref="GI:28872"
                  /db_xref="SWISS-PROT:P00966"
                  /translation="MSKSGSVLAISGGLDTSCILWLKEGYDIYAVIANTIQKEDF
VEIARQKALIKGAKKVI LEDVSRFEVEFWPAILQSSALYEDRLTGLTSLARPIARKO
METAKQKQIRIPVTPKPMMSDENLMIHISEAGILLENPKNOAPDGLYTTODPARAPN
TPDLIEIEFKKGVVYKVDGTTHOTSLLEMYENVEAGHVGGRIDIVENRRTGM
KSRGIETTPAGTILYHAHLDIEAFITMDREVRIKQSLGKPAELIYVYTGIRSPCEEFV
RHCIASQOEVEKVOYSVIKQGVYILGRESPLSTINEELVSMNVQGDYEPTDGTGFI
NINSLRKEVHRIOKSVYTK"
variation         759..760
                  /note="pot. additional A in PAS1"
variation         1320
                  /note="U is C in variant PAS2"
variation         1431
                  /note="G is A in variant PAS1"
misc_feature      1526..1532
                  /note="put. polyadenylation signal"
misc_feature      1537..1544
                  /note="pot. polyadenylation signal"
polyA_site        1547
                  /note="polyadenylation site"
variation         1555
                  /note="A is U in variant PAS2"
BASE COUNT       393 a      431 c      434 g      302 t
ORIGIN
Query Match      100.0%; Score 1239; DB 9; Length 1560;
Best Local Similarity 100.0%; Pred. No. 4,1e-247;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 541 atgataagaacatcatatgacatcagctacgagctctggaatccttgagaacccccagaac 600
    |||||
DB 616 ATGGATGAGAACCTCATCATCATCAGCTACGAGGCTGGAATCCTCGAGAACCCCAAGAAC 675
OY 601 caagcgctcctcaggctctacacgaagaccgaagaccgaagcccccaacacccct 660
    |||||
DB 676 CAAGCGCTCCAGGCTCTACACGAGACCCAGACCACCAAGCCCCCAACACCCCT 735
OY 661 gacattctcagatcagatctcaaaaaggggtccctgtgaaggtagcaacgtcaagat 720
    |||||
DB 736 GACATCTCGAGATCGAGTTCAAAAAAGGGGTCCCTGTGAAGGTGACCAACGTCAGAGAT 795
OY 721 ggcacaccccccaagacctcctctgagctctcatctatctacatgaagaatctcgggcaag 780
    |||||
DB 796 GGACACACCCACGAGACTCTCTTGAGCTTTATGTACTGACGAGAACTCCGGGCAAG 855
OY 781 catggctgggctgctatgacatcgttgagaacacctctatgtgaatgaagtcctcgaagt 840
    |||||
DB 856 CATGGCTGGGCGCTTATTGACATCGAGAGAACCGCTTCATTGGAATGAAGTCCGAGGT 915
OY 841 atcaagagacccccagagaccacatccttaccatgctcattgacatcgaagctctc 900
    |||||
DB 916 ATCTACGAGACCCAGCAGCAGCACCACCTTTACCATGCTCATTTAGACATCGAGGCTTTC 975
OY 901 accatgaccgggaagtgcgcaaatcaacaagaagccttgagcttgaaatttgctgagctg 960
    |||||
DB 976 ACCATGACCCGGGAGAGAGTCCCAAAATCAAAAGGCGCTGGCTTGAATTTCTGAGCTG 1035
OY 961 gttatcacgggtttagcgctgacctgagctgtgaattgtctgcacatgcacgaag 1020
    |||||
DB 1036 GTGATACCGGTTTACGAGGCTTACGCTTGAAGTGTGATTTGTCCGACCATGCTGCCAAG 1095
OY 1021 tcccaagagcgaatggaaggaagagtgacagctgctcctcaagggcgaagtgtaact 1080
    |||||
DB 1096 TCCAGAGAGCGAGTGAAGGAGAAAGTGCAGGTGCTGCTCAAGGCGCAGGTGACTTC 1155
OY 1081 ctggcgccgggaagctcccatctctctctacaaatgagagctgtgtgagaatgaagctgca 1140
    |||||
DB 1156 CTGGCGCGGAGTCCCACTGTCTCTTAACAATGAGGAGTGTGACATGAAGCTGCAG 1215
OY 1141 ggtatlatgaagcaactatgacacgggttaccatcaatctccctgaagctgaag 1200
    |||||
DB 1216 GGTGATTATGAGCAACTGATGCCACCGGTTTCATCAATCAATTCCTTCAGGCTGAAG 1275
OY 1201 gaatatcatgctctccagaagaagctcaatgccaatag 1239
    |||||
DB 1276 GAATATCATGCTCTCAGAGCAAGGTCACTGCCAATATG 1314

RESULT 2
AK027126      1975 bp      mRNA      PRI      29-SEP-2000
LOCUS        Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
DEFINITION   HSASD Human mRNA for argininosuccinate synthetase.
ACCESSION   AK027126
VERSION      AK027126.1 GI:10440175
KEYWORDS     oligo cloning; fis (full insert sequence).
SOURCE       Homo sapiens human small intestine cDNA to mRNA, clone_11b:HSI
             clone:HS113532.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (sites)
AUTHORS     Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
             Shibahara,T., Tanaka,T., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
             NEDO human cDNA sequencing project
             Unpublished (2000)
TITLE       2 (bases 1 to 1975)
REFERENCE   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
             Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE       Direct Submission

```

Med Feb 13 07:36:01 2002

JOURNAL

Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT

NEO human cDNA sequencing project supported by Ministry of International Trade and Industry for Japan: cDNA full insert sequencing. Research Association for Biotechnology: cDNA library construction. 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES

SOURCE

Location/Qualifiers
1. 1975
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HS11352"
/clone_lib="HS1"
/issue_type="human small intestine"
/note="Cloning vector pME18SFL3"

misc-feature
1. 1975
/note="highly similar to HSASD Human mRNA for argininosuccinate synthetase"

BASE COUNT

459 a 564 c 563 g 389 t

ORIGIN

Query Match 99.5%; Score 1232.6; DB 9; Length 1975;
Best Local Similarity 99.7%; Pred. No. 8.4e-246;
Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 atgttcagcaaaagctcctggtgtctgagctacagtgccgctcggagcactcgtgcatc 60
|||||
482 ATGTCCACCAAGGCTCGTGGTTCGGCTACAGTGGGCGCTGGACACCTCGTGCATC 541

61 ctctgtgtgctgaagaaagcgtatgagctatgctatcgtgccaacattggccag 120
|||||
542 CTGTGTGGCTGAAGAACAGGCTATGACGTCATTGCCATCTGGCAACATTTGGCCAG 601

121 aaggaagacttcgagaaagccaggaaggaagcactgaagcttggggccaagaagtgctc 180
|||||
602 AAGGAAGACTTCGAGAACCCAGGAAGGCAAGCTGAAGCTTGGGGCCAAGAAAGTGCTC 661

181 attgaagatgctcagcaggaagtttggagagatcactcgtgcgcgactcagtcagc 240
|||||
662 ATTGAGCATGTCAACAGGAGTTTGGAGAGTTCAATCTGGCGGCATCCACAGTCACAGC 721

241 gacatgtatgagacgctactcctctggtgacactctcttgcagagccctgcatcgccg 300
|||||
722 GCACTGTATGAGACCGCTACTCTCTGGGACCTCTCTTGGCCAGGCCCTGCATCGCCGC 781

301 aaacaaagtgaatcgcgcagcagcaggaaggggccaagtatgtgtccacagcgccaca 360
|||||
782 AAACAAGTGAATCCGCCAGCGGAGGGGCCAAGATGTGTCCACAGGGCCGCCACAGA 841

361 aaggggaagcatcagtcggtttgagctcagctgctactactctgccccccagaataag 420
|||||
842 AAGGGGAACGATCAGTCCGGTTTGAAGCTCAGCTACCTACTGCGCCGCCAGTAAGA 901

421 gtctgtcctccctgagagatgctcgaattcttacaacccggttcaagggcgcaatgacct 480
|||||
902 GTCATTCCTCTCGAGAGATGCTGTAATCTCAACCGGTTCAAGGGCCGCCATATGACCTG 961

481 atgagctcgaagcaaacacggatcccatcccggtcactcccaagaacccgtgagc 540
|||||
962 ATGAGCTCGAAGCAAAACAGGGATTCCTCCGCTACTCCCAAGAACCCCTGTGAGC 1021

541 atgagtgagacatcagctacagctcagagctggaatcccttgaaagaacccaagaac 600
|||||
1022 ATGATGAGAACCTCATGACATCAGCTACAGAGGCTGGAACTCTGGAGAACCCCAAGAC 1081

601 caagcgctccagtgcttacaagaagccaggaaccagcaagaagccccaacacccct 660

|||||
1082 CAGGCGCTCCAGCTCTCTACAGAGACCCAGACCAGCCAGCAAGCCCAACACCTC 1141

661 gacattctgagatcaggttccaaaagggctctgtgaaggtgagcaagctcaagat 720
|||||
1142 GACATTCGGAATCGAGTTCAAAAAGGGGCTCCGTGAAGGTACCAACGTCMAAGAT 780

721 ggcacacccacacgaactcctcttgagctcttactgacctgaagaaggtcgagggaag 780
|||||
1202 GGCACACCCACACGACCTCTTGAGAGTCTTCAATGACCTGAACAGCTCCGGGCAAG 1261

781 catggtgtggccgtatgtacatgtggaagacccgttcaattggaatgaaagtcagag 840
|||||
1262 CATGGCTGGGGCGTGTGACATCTGGAGAACCGCTTCAATGGAATGAATGCCAGGT 1321

841 actacagagacccagcagagcaccatcttaccatgctcatctttagacatcagagcttc 900
|||||
1322 ATCTACGAGACCCAGACGACGACCATCTTTACATGCTATTAGCATTCGAGGCTTC 1381

901 accatgagacccggaagtgcgcaaaatcaaacagccttgagcttgaatttgcagatc 960
|||||
1382 ACCATGAGACCCGGAAGTCCCAAAATCAAAACAGGCTGGCTTGAATTTGCTGAGCTG 1441

961 ggtataccggtttacagcagccagccctgagatgtgaaattgtcgcactgcatcgcaag 1020
|||||
1442 GTGTATACCGTTTCTCGACAGCCCTGAGTGAATTTCTCCGCACTGATTCGCCAAG 1501

1021 tcccaagagcagtggaaggaagtgcaagtgctcgtcctcaagggcaggtgtacatc 1080
|||||
1502 TCCAGAGGCGAGTGGAGAGGAGAAAGTGAGTGTCCGCTCAAGGCCAGGTGTACATC 1561

1081 ctggcggagagtcctccactgtctcttcaaatgagagctgtgagcatgacgtgcag 1140
|||||
1562 CTGGCGGAGTCCCACTGTCTCTTACATGAGGAGCTGTGAGCATGAAAGTGCAG 1621

1141 ggtgattatgagcaacatgagtcacacgaggttcaataatcaattccctcaagctgaag 1200
|||||
1622 GTGTATATGAGCAACATGATGCCACCGGTTCATCAATCAATTCCTCAGCTGAAG 1681

1201 gaatatcatgcttccagagcaaggtacatcagccaatag 1239
|||||
1682 GAATATCATGCTCTCCAGAGCATGCTCCAAATAG 1720

RESULT 3
BC009243 1595 bp mRNA PRI 12-JUL-2001

LOCUS Homo sapiens argininosuccinate synthetase, clone MGC:3634
DEFINITION IMAGE:3010137, mRNA, complete cds.

ACCESSION BC009243
VERSION BC009243.1 GI:14328058

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

Strasbourg, R.
Direct Submission
Submitted (06-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.W.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadan@systemsbio.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia

Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INTL at: <http://image.llnl.gov>
Series: IRAL Plate: 10 Row: e Column: 18
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4537336.

FEATURES

source

Location/Qualifiers

1. 1595

/organism="Homo sapiens"

/db_xref="taxon:9606"

/db_xref="GI:14328053"

/issue="MGC:3634 IMAGE:3010137"

/clone_type="Muscle, rhadomyosarcoma"

/lab_host="NIH-MGC-17"

/notes="Vector: pOTB7"

81. 1319

/codon_start=1

/product="argininosuccinate synthetase"

/protein_id="AAH09243.1"

/translation="MSSKGVSVLAISGLDTSIIYWKKEGYDIVIYLANIGQEDF

VEIAORBAKAYVSHGATGKNDQRELSYLAQIVIAIPMPPEYRNRKQEDF

TPDIETIEPKGVKVPKPMMDENLHISYEGLIENKRNAPGGLYKTODAPKRN

KSRGIEPFGKGVKVPKPMMDENLHISYEGLIENKRNAPGGLYKTODAPKRN

RHCTAKSOVEKGVKVPKPMMDENLHISYEGLIENKRNAPGGLYKTODAPKRN

NINSLRKEVHRISKVTAK" 440 g 312 t

BASE COUNT

400 a 443 c 440 g 312 t

Query Match

Best Local

Similarity 99.4%; Score 1231; DB 9; Length 1595;

Matches 1234; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 atgtccagaagaagctcgtgttctgagcctacagtgagcgagcctgagcctcgtgacac

81 atgtccagaagaagctcgtgttctgagcctacagtgagcgagcctgagcctcgtgacac

61 ctgtgtgtgtgaggaagaagcctatgagcctatgagcctatgagcctatgagcctatgagc

141 ctgtgtgtgtgaggaagaagcctatgagcctatgagcctatgagcctatgagcctatgagc

121 aaggaagacttcgaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga

201 aaggaagacttcgaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga

181 atgtgagatgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

261 atgtgagatgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

241 gcaatgatgaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

321 gcaatgatgaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

301 aagaagtggaatcgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

381 aagaagtggaatcgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

361 aagaagtggaatcgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

441 aaggggaacgagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

421 gtcatttccttccttcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

501 gtcatttccttccttcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

481 atgtgagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

561 atgtgagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

TITLE
JOURNAL
COMMENT

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lakoque, K., Lamazares, R., Landers, J., Lenoczky, J.,
Levine, R., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., McMan, P., McGuck, A., McKernan, K., McPheters, R.,
Meidrim, J., Menus, L., Mihova, C., Miranda, C., O'Donnell, P.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: U5602
Center clone name: 541_O_3

Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158974 bases at least Q40
Consensus quality: 169058 bases at least Q30
Consensus quality: 174233 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 178138; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 28 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 315: contig of 315 bp in length
316 415: gap of 100 bp
416 1503: contig of 1088 bp in length
1504 1603: gap of 100 bp
1604 3415: contig of 1812 bp in length
3416 3515: gap of 100 bp
3516 5368: contig of 1833 bp in length
5369 5468: gap of 100 bp
5469 7164: contig of 1696 bp in length
7165 7264: gap of 100 bp
7265 9123: contig of 1859 bp in length
9124 9223: gap of 100 bp
9224 10641: contig of 1418 bp in length
10642 10741: gap of 100 bp
10742 11946: contig of 1205 bp in length
11947 12046: gap of 100 bp
12047 14424: contig of 2378 bp in length
14425 14524: gap of 100 bp
14525 17721: contig of 3197 bp in length
17722 17821: gap of 100 bp
17822 20638: contig of 2817 bp in length
20639 20738: gap of 100 bp
20739 24567: contig of 3829 bp in length
24568 24667: gap of 100 bp

FEATURES

source

24668 29147: contig of 4480 bp in length
29148 29247: gap of 100 bp
29248 33430: contig of 4183 bp in length
33431 33530: gap of 100 bp
33531 38145: contig of 4615 bp in length
38146 38245: gap of 100 bp
38246 43517: contig of 5272 bp in length
43518 43618: gap of 100 bp
43618 48556: contig of 4939 bp in length
48557 48656: gap of 100 bp
48657 55104: contig of 6448 bp in length
55105 55204: gap of 100 bp
55205 55983: contig of 4629 bp in length
55984 59933: gap of 100 bp
59934 66011: contig of 6078 bp in length
66012 66112: gap of 100 bp
66112 72673: contig of 6562 bp in length
72674 72773: gap of 100 bp
72774 79925: contig of 7157 bp in length
79926 80025: gap of 100 bp
80026 85367: contig of 5342 bp in length
85368 85467: gap of 100 bp
85468 92616: contig of 7149 bp in length
92617 92716: gap of 100 bp
92717 101894: contig of 9178 bp in length
101895 101994: gap of 100 bp
101995 118357: contig of 16363 bp in length
118358 118457: gap of 100 bp
118458 138888: contig of 20431 bp in length
138889 138988: gap of 100 bp
138989 180838: contig of 41850 bp in length.

Location/Qualifiers

1. 180838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-54103"
/clone_1bp="RPCT-11 Human Male BAC"
1. 315
/note="assembly-fragment"
clone_end:17
vector_side:left
416 1503
/note="assembly-fragment"
1604 3415
/note="assembly-fragment"
3516 5368
/note="assembly-fragment"
5469 7164
/note="assembly-fragment"
7265 9123
/note="assembly-fragment"
9224 10641
/note="assembly-fragment"
10742 11946
/note="assembly-fragment"
clone_end:sp6
vector_side:right
12047 14424
/note="assembly-fragment"
14525 17721
/note="assembly-fragment"
17822 20638
/note="assembly-fragment"
20739 24567
/note="assembly-fragment"
24668 29147
/note="assembly-fragment"
29248 33430
/note="assembly-fragment"
33531 38145
/note="assembly-fragment"

Query Match	88.2%	Score 1093;	DB 2;	Length 180838;
Best Local Similarity	93.7%	Pred. No. 3.4e-217;		
Matches 1161; Conservative	0;	Mismatches 75;	Indels 3;	Gaps 2;

QY	601	caaggccctccagaggtctctaaacgaagcccgagcccaagcccaagcccccaaacccct	660	
Db	147731	CAAGCACTCCAGGTCTCTACACGAAGA	CCAGGACCGGGCCAAAGCCCCCAACACCTT	147783
QY	661	gacatctctgagatcogagttcaaaaaaggggtccctgtgaaggtgaacaaagtcgaagat	720	
Db	147790	GACATTCTTGAAATGAGATTCMAAAAAAGAGTCCCTGTGAAGGTGCACCAAGCTCAAGAT	147845	
QY	721	ggcacaccacacgaacacctcccttgagctctctcaatgaacctgaaacgaagtcggggcaag	780	
Db	147850	GGCACACCACCCAGACCGTCATTTGGAGCTCTTCATGTACTGAACGAAGTGTGGGCAAG	147905	
QY	781	catggcgtggggcgcgatattgaacatcgtygaagacogctcattcgtgaatgaagtcceagat	840	
Db	147910	CAGGGTGTGGGCTGTATTGACATCTATGGAAGAACCGCTTCATTTGAATGAAGTCCCGAGG	147965	
QY	841	alctacgaagacccacagcagcacatcccttaccatgtcattagaacacgaagcctc	900	
Db	147970	ATCTACGAGGCCCCGACGACCAACATCCTTTACCTGGCTATTAGACACCAAGGCTTC	148029	
QY	901	accatggagccgggaagtcgcgaatatcaaaaagcctggagcttgaattgtcgtagctg	960	
Db	148030	ACCATGGGCGCAGGAAGTCCGAATTCAAACAAGGCGCTTGAGAAATTTGTTGAGCTG	148089	
QY	961	gtgtatccoggtttacggtctagccctgaagttgaatttgcgcgaactgcacgcgaag	1020	
Db	148090	GTGGATTACCGGTTCTCGGTACAGCCCTGAGGTGTAATTTGTCGCCACTGCATCGCCAG	148149	
QY	1021	tcccagagagcagatggaaaggaagatgcaggtgcgcgtccctcaagggccaggtgtacatc	1080	
Db	148150	TCCAGGAGCAAGTGGAAAGGAAGTCCAGGTGTCCTCTCGAGGGCCAGGTGTCAATC	148209	
QY	1081	ctcggcgcgggaagtcccaactgtctctctacaatgaagagcttgtylgacalgaacgtgcag	1140	
Db	148210	CTTGCCCAAGGAGTCCCACTGTCCTCTACAAAGAGGAGCGCGGTGAGCAAGACGTGCAG	148269	
QY	1141	ggtyattatggagccaactgatgtcacacgggttcatacaactcaatccctccagctgaag	1200	
Db	148270	GGTGAATTATCAGCAATTGATGTCGACCGGGTTATCAACACTCAATTCCTCTCAGGCTGAAG	148329	
QY	1201	gaatatcatcgtctccgaagcaaggtctactgcgcaaatag	1239	
Db	148330	GAAATATCATCTCTCCAGACCAAGGTCTACGTCCCAAAATAG	148368	

```

RESULT      5
AC025175/c
LOCUS
DEFINITION
AC025175      165439 bp      DNA      HTG      18-JUL-2000
Homo sapiens chromosome 5 clone CTD-2081C10, WORKING DRAFT
SEQUENCE      20 ordered pieces.
AC025175
AC025175.3      GI:9256455
HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (07-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7711790.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----

```

```
Project Information
Center Project Name: 658711
Center clone name: CITB-H1_2081C10
```

Summary Statistics

* NOTE: This is a 'working draft' sequence. It

* consists of 20 configs. Gaps between the configs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

*	75	874:	contig of 174 bp in length
*	875	21897:	contig of 21023 bp in length
*	21898	21997:	gap of unknown length
*	21998	26272:	contig of 4275 bp in length
*	26273	26372:	gap of unknown length
*	26373	60196:	contig of 33824 bp in length
*	60197	60296:	gap of unknown length
*	60297	68062:	contig of 776 bp in length
*	68063	68162:	gap of unknown length
*	68163	78032:	contig of 9870 bp in length
*	78033	78132:	gap of unknown length
*	78133	82251:	contig of 4119 bp in length
*	82252	82352:	gap of unknown length
*	82352	99537:	contig of 17186 bp in length
*	99537	99637:	gap of unknown length
*	99638	105678:	contig of 6041 bp in length
*	105679	105778:	gap of unknown length
*	105779	107011:	contig of 1239 bp in length
*	107018	107117:	gap of unknown length
*	107118	108655:	contig of 1538 bp in length
*	108656	108755:	gap of unknown length
*	108756	115076:	contig of 6321 bp in length
*	115077	115176:	gap of unknown length
*	115177	117477:	contig of 2301 bp in length
*	117478	117577:	gap of unknown length
*	117578	136876:	contig of 19299 bp in length
*	136877	136976:	gap of unknown length
*	136977	140794:	contig of 3818 bp in length
*	140795	140894:	gap of unknown length
*	140895	154782:	contig of 13888 bp in length
*	154783	154882:	gap of unknown length
*	154883	155882:	gap of unknown length
*	155883	156598:	contig of 1716 bp in length
*	156599	161013:	contig of 4315 bp in length
*	161014	161113:	gap of unknown length
*	161114	162823:	contig of 1710 bp in length
*	162824	162923:	gap of unknown length
*	162924	165493:	contig of 2516 bp in length.

```

FEATURES
  source
    Location/Qualifiers
      1..165439
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="CTD-2081C10"
        /clone_lib="caltech human BAC library D"
BASE COUNT
  45132 a 3811 c 33426 g 52152 t 1918 other
ORIGIN

```

Query Match	Score	DB 2:	length
Best Local Similarity	88.1%; 93.6%;	1091.4; Pred. No. 7.3e-217;	165439

Matches 1160; Conservative 0; Mismatches 76; Indels 3; Gaps 2.

QY	1	atgcgcgaagaagcccgcttgctctcgtccaaagctgaggcgcttgacaacctgctgaac	60
Db	52611	ATGTCACCAAAAGCCTCCCTGGTTCTTGACCTACAGTGCGACGCTGGACACCTCCCTGCACCC	52555
QY	61	ctcgtctgctgaagaacaaagcctatagaactgaactgttcctatctgcgcaaatgtgcac	120
Db	52551	CTCCTGTGGCTTAAGAAACAGAGCGCTATGACGTACATTGGCTACCTGGCCACACTTGGCCAG	52495
QY	121	aaggaagactctgaagaagcaagaagaagacctaagaactgtgggccaanaagtgctc	180
Db	52491	AAGGAAGCTTGTGAGGAAGCCACGAAAGATGCACGTGAATTTGGGCTTAAGAAAGGTTC	52435
QY	181	attgaagatgtcaacgaaggaattgtggaaggaattcatctgacggccatccagctcagc	240
Db	52431	ATTGAAGAAATGTGGGACGAGGAGTTGTGTGAGAGATCATCTGCGCCGGCCATCACTGACGT	52375
QY	241	gcaacttatgaagacgcgtactcctcctgtggacctctctgcaagccctgtatcgcgcgc	300
Db	52371	GCACGTATGAGAGACCGCTGACTCTTGTGGACACTCTC--GCGAGGCGCTGTATGCGCTGC	52314
QY	301	aaaacaatgfgaaatcgcgccagcggagaggggccaagatagtgtgtccacgagcgccacaaga	360
Db	52313	AAACAATATGGAATACCCCACTGGAAAGGGGGCCAGTATATGTCCACAGCAGCACGGGA	52255
QY	361	aagggaagacgatcaggttcgcggtttgagctcgaactctactactacgtgcccccaataag	420
Db	52253	AATGGGAACATTCAGAGTCCGTTTGAGCTCATTTTGTCTACTCACTAGGCCCCCGAGTAAG	52194
QY	421	gtccattgtctccctgaggagatgacctgaattctacaacccggttccaaggcgccaatgacctg	480
Db	52193	GTCTATTGCTCCCTGGAGAAATGCCAAATTCATCAACACAGCTCAAGGGCCGCAATGACCTG	52134
QY	481	atcgagctacgcaagaacacacaggaattccaccacccggtcactcccaagaacccgtgagc	540
Db	52133	ATGGAATATGCAAGCAACACAGAGATTTCCCATCCGGCTACACCCAAAGAACTGTGTGAC	52074
QY	541	atggatgtgaactcctcatgtcaatcaactcaagagcttgaactcgtgaaaccccaagaac	600
Db	52073	ATGGACGGAACCTCAAGCATATACCTACCTACGAGTCTGGAAATCTCGAGAACCTTAAGAAC	52014
QY	601	caagcgctcccaagctctctacaagaagaccagagcccaagccaaagccccaacccct	660
Db	52013	CAAGCACTCCAGGTCTCTACACGAAGA--CCAGAACCCGGCCAAAGCCCCCAACACCTCT	51955
QY	661	gaactcttgagatctgaattcaaaaaaggggtccctgtgaaggttgaaccaagtcaagga	720
Db	51954	GACATTCTCAGAGATCAGATTCAAAAAAGAGATCCCGTGAAGGTGACCAACGTCAAAGAT	51895
QY	721	ggcacaccaccacagacccctcctcttgagctctctctactgaagaagaatctgcgggaag	780
Db	51894	GGCACACCCACACAGACTCATATTGAGGCTCTTCACTGATACCAAGCAAGTCTGTGGCAAG	51835
QY	781	catgggtgggcgcgttatgaactcgttggagaacgcgttcaatgtgaatgaagtcgccgaagt	840
Db	51834	CAGGGTGTGGGCTGTATTTACATCATGTGAGAACCGCTTCTATTGGAATGAAGTCCGGAAGG	51775
QY	841	atctacgaagcccccagcagagcaaccttctacacactctatctaatgaagaatgaagccctc	900
Db	51774	ATTACACAGAGCCCAACAGACACCATCTTACTCTGCTCATTTTGAACACCAAGGCTTTC	51715
QY	901	accattgaccgcgggaagtgcgcataaatacaacaagcctgtggcttgaatttgcgtgagctg	960
Db	51714	ACCATGGGCCAGGAAGGTGCGCAAAATCAACACAGGCCCTGGCGCTGAAAATTTGTGACTCG	51655
QY	961	gtgtataccggtttaaaccgctcagccctgaatgttgaatttgcctgcgcactgatctgccaag	1020
Db	51654	GTCGATACCGGTTCTTCGGACACGGCCGTGAGTGTGTTGTCTGCCACTGCATGCCAAG	51595
QY	1021	tcccagaagccggatggaaaggaagctgcaggtctccgtctccaaggcgcaagtgtatcatc	1080
Db	51594	TCCCAAGACCAAGTGAAGGGAAGGTCCAGTGTCCGCTCTCGAAGGCCCAAGTTCATATC	51535

Dp	190	ATTGAGGACATCGAGAGAGAGTTTGAGAGAGATTTCATCTGGCGGGCCATCCAGTCCAGC	249
Oy	241	gcactgtatgaagacccgtactcctcttgggaacctctctgccaagcccttgatgcgcgc	300
Dp	250	GCACTGTACGAGGACCGATACCTCTCTGGGACCTCTCTGCGCAGGCCCTCGATCCCGCG	309
Oy	301	aaacaagtgaatatcgcccccgggggggggccaatatgtgtcccaagcgcccaaga	360
Dp	310	AAGCAGGCTGGAGATCGCCACGAGCAAGGAGCCAAATATGTGTCTCACGGCCGCCACAG	369
Oy	361	aagggaagcaatcaagtcgcggttttgagctcagcttgctactcaatctgccccagaataag	420
Dp	370	AAGGGGAAACGACCGATCCGGTTTGAGCTTCACTCGTACTCGGTGGCCCGACAGATCAAG	429
Oy	421	gtcatctgcctctggagagatgtcctgaattctacaaccggttcaagggccgcaatcctg	480
Dp	430	GTCACTGCTCCCTGGAGGATGCCGAGTTCTTATACCGGCTTCCAGGGCCGCAACGATCTG	489
Oy	481	atggaatgcgaagaagcaaacacggattcccatcccggtcactcccaagaaccgttgagc	540
Dp	490	ATGGAGTATGGGAAGCAACATGGATCCCGCTCCAGTACACCCGCCAAGAACCGGTGAC	549
Oy	541	atggaatgaagacctcatgacatcagctacagagcttgaaatcctctggagaacccccagaac	600
Dp	550	ATGGACGAGAACTGATTCATATACGTACGAGGCTGGAATCCTGGAGAACCCCAAGAAC	609
Oy	601	caagcgctccaaagtctctacagaagaccgaagaccgaacgaagcccccaaacacct	660
Dp	610	CAGGCGCTCCAGGCTCTTACACAAAGCCAGGACCCGGCCAAAGCCCCAACACCCG	669
Oy	661	gacattctcgaatcagattcaaaaagaggctccctgtgaagttgaccaaagtcaagat	720
Dp	670	GACATCTCTGAGATGAGATTCAAGAAAGGGTCCCGTGAAGGAGACCAACGTCTGGGAGAT	729
Oy	721	ggcaccaccacacgaacctcctcttggagctctcatctacactgaacgaatcgggggcaag	780
Dp	730	GGCACCACCCACAGACGACCGTTGGAGCTTTCTCTGTAACGAATGAAGTGGCTGGCAAG	789
Oy	781	catgctgggcccgtatgaactctgttgagaacccctcatcttgaaatgaagtcocgaagt	840
Dp	790	CAGGCGGTGGGCGCATCGACATCTGGAAACCCGCTTCACTGGGATGAAGTCCCGGGGT	849
Oy	841	atctacgaagaccacgaagcagacacacacctcttacaatgtcatcttgaacatcagagcctc	900
Dp	850	ATCTACGAGACCCGACGGGAGCGAGATCTTACACGCTCATTTAGACATCGAGGCTTC	909
Oy	901	accatggaaccgggaagtgcgcgaataatcaaacaaagccgggcttgaaattgtctgagctg	960
Dp	910	ACCATGAGACCGGGAATGTGGCAAAATCAAGCAAGGCTCTGGGTTGAATTGGCCGAGCTG	969
Oy	961	gtgtataccggtttacgacctagccctgagtgtgaattgtctcgcacatgcatgcgaag	1020
Dp	970	GTCTACACGGGTTTCTGCGCACAGCCCGAGTGTGAATTGTCCGCATCTGATTCGCAAG	1029
Oy	1021	tcccaagagcgagttggaagggaagtgaagtgctcgtctcctaagggccaaggttlaatc	1080
Dp	1030	TCCAGAGAGCGCTGGAAGGGAAGTGCAGGTGTCCGTTCTAAAGGCCAGGTGATCTC	1089
Oy	1081	ctcgccgggaagtccccaactgtctctctacaatgaggaagctgtgtagacatgaaagtgcag	1140
Dp	1090	CTTGTGGCGGAGGTCCCATCTGCTCTCTCAATGAGAGACTGTGAGCATGAACGTGCAG	1149
Oy	1141	ggtgaattatgagccaactgtatgcaccgggttcataacaatcaatctcctcagctgtaag	1200
Dp	1150	GGAGACATACGAGCGGGTTGATGTGCACACGGTTTCATACATCAATTCCTCAGGCTGAAG	1209
Oy	1201	gaatatcatcgtctccaagagcaaggttcactgccaatatg	1239
Dp	1210	GAATATCATCCCTCCAGAACCAAGGTCTACCCGCAAAATAG	1248

AC009550
 LOCUS AC009550 188356 bp DNA HTG 20-APR-2000
 DEFINITION Homo sapiens clone RP11-382M14, WORKING DRAFT SEQUENCE, 11
 AC009550
 ACCESSION AC009550.3 GI:7622313
 VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome, clone RP11-382M14
 JOURNAL Unpublished
 AUTHORS 2 (bases 1 to 188356)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
 Marquis, N., McEwan, P., McEwen, A., McKernan, K., McLoughlin, J.,
 Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Teefaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 DIRECT SUBMISSION
 Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 20, 2000 this sequence version replaced gi:6006236.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 COMMENT
 TITLE JOURNAL
 JOURNAL
 COMMENT
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center Project name: L1662
 Center clone name: 382_M_14
 Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-primer-amersham; 4% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 159465 bases at least Q40
 Consensus quality: 175637 bases at least Q30
 Consensus quality: 182085 bases at least Q20
 Insert size: 194000; agarose-fp
 Insert size: 187356; sum-of-contrigs
 Quality coverage: 4.4 in Q20 bases.
 Quality coverage: 4.6 in Q20 bases.
 NOTE: This is a 'working draft' sequence. It currently
 consists of 11 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1 1698: contig of 1698 bp in length
 * 1698 1798: gap of 100 bp
 * 1799 4670: contig of 2872 bp in length
 * 4671 4770: gap of 100 bp
 * 4771 8967: contig of 4197 bp in length
 * 8968 9067: gap of 100 bp
 * 9068 13012: contig of 3945 bp in length

13013 13112: gap of 100 bp
 * 13113 21156: contig of 8044 bp in length
 * 21157 21256: gap of 100 bp
 * 21257 29479: contig of 8223 bp in length
 * 29480 29579: gap of 100 bp
 * 29580 48365: contig of 18786 bp in length
 * 48365 48465: gap of 100 bp
 * 48465 67005: contig of 18540 bp in length
 * 67005 67105: gap of 100 bp
 * 67105 89097: contig of 21992 bp in length
 * 89098 89197: gap of 100 bp
 * 89198 117364: contig of 28167 bp in length
 * 117365 117464: gap of 100 bp
 * 117465 188356: contig of 70892 bp in length.
 FEATURES
 source
 1. 188356
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-382M14"
 /clone_lib="RPCT-11 Human Male BAC"
 1. 1698
 /note="assembly-fragment"
 1798. 4670
 /note="assembly-fragment"
 4771. 8967
 /note="assembly-fragment"
 9068. 13012
 /note="assembly-fragment"
 clone_end:sp6
 vector_side:left
 13113. 21156
 /note="assembly-fragment"
 21257. 29479
 /note="assembly-fragment"
 29580. 48365
 /note="assembly-fragment"
 48465. 67005
 /note="assembly-fragment"
 clone_end:t7
 vector_side:left
 67106. 89097
 /note="assembly-fragment"
 89198. 117364
 /note="assembly-fragment"
 117465. 188356
 /note="assembly-fragment"
 BASE COUNT 57641 a 37359 c 35201 g 57123 t 1032 others
 ORIGIN
 Query Match 85.4% Score 1058; DB 2; Length 188356;
 Best Local Similarity 92.2%; Pred. No. 6,1e-210.
 Matches 1151; Conservative 0; Mismatches 85; Indels 12; Gaps 3;
 QY 1 atgtcagcaaaagctcgtggtctgtgctctacagctgctgctgcaacccgtgacac 60
 Db 121401 ATGTCCAGCAAAAGCGCTCTATGCTTGGCGTGCAGCGACGCTGACACCTTGGCATT 121460
 QY 61 ctgtgtgtgtgctgaagaaacaaagctatgacgtcattgctctatctgtgcaacattggcag 120
 Db 121461 CTGTGTGTGCTGAAGAAAGCGCTATGACGTCTGCTTGGCTGCGCAGCATTTGGCCAG 121520
 QY 121 aaggaagacttcagaaagcaagcaagaaagcaagcaagcaagcttggggccaaagagtggttc 180
 Db 121521 AAGGAAGACTTCAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 121580
 QY 181 attgaagatgctcagcagcagagcttctgtgagagagcttcatctggtcgcagccatccagtcagc 240
 Db 121581 ATTGAAGATGCTCAACAGGAGATTGTGGAGGAGTTCATCTGGCCGCAATGCATGCACAGC 121640
 QY 241 gcactgtatgagagcagcgtctactctctgtggcaacctctctgtcagagcctgtcagtcagccgc 300
 Db 121641 ACACCTGTATGAGAGACGCTACCTCTGTGGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 121700

Chr1	301	aacacatgagaaatccgcacgagcgggaaggggccaagtctgtccaaagcgccagaa	360
Chr1	121701	AAACAAGTGGAAAGTCGCCACCGGAGGGGGCCAAATATTGTCTCCACAGGCACACGGGA	121760
Chr1	361	aaggagaaacgaatcaggtccggtttgaagtcagctgtactcaactgtgccc-cagataaa	419
Chr1	121761	AAGGGAGATGATGATCAGGTGTGGTTTGAGCTCACCAGTGCCTACTGCTGGGCCCCACAGATAAA	121820
Chr1	420	ggtcattgtctccctcggagatgtccgtgaattttaaaccggtcttaaggcccgcaatgacct	479
Chr1	121821	GGCTATTGCTCTCCATGAAGATATAGCCCGAATTTCTACAAACGGTTCAAGGGCCACAGTAGACTT	121880
Chr1	480	gatggaagcgaacgaagcaacacggagatctccatcccggtctactcccaagaacccgttgag	539
Chr1	121881	GACGGAATATATGCAMAAGCAACCGGGATTTCCACCCCGGTTCACCTCCACGAAGAACCGTGGAG	121940
Chr1	540	catgatatgaagaacctcatgacatctacgtacagagcttgaatcctgtgaagacccaaga	599
Chr1	121941	CATGGACGAGAAACCTCATAGCAATCAGCTGGAGGCTCGAATTCGGAGAAATCCCGAAGAT	122000
Chr1	600	ccaagcgctccaggtctctacaagaaaccagaaccgaacccaagaagccccaacacc	659
Chr1	122001	CCAAACGCTCCAGCTCTCTACACGAACAACCCAGGACCCGGCCAAACCCCAACACCC	122060
Chr1	660	tgaactctcagatcagatctcaaaaagggttcccttgaagtgtgacaaagtcaagaa	719
Chr1	122061	TGACATTCTTCGAGATCGAGTTCAAAAAGGGGTCCCGCTGAAGGTGACCAACGTCAGAGA	122120
Chr1	720	tggcaaccaaccaagacctctctgtgagctctctcatgtactgtgaacgaagtccggtgcaa	779
Chr1	122121	TAGTGTCCACCCACACAGACCTCTCTTGAGCTCTCTCATGTATTCGAACGAAGTGGCGGAA	122180
Chr1	780	gcatggcgttgagcgcgtattgacatctgtgagaaaccgctcatctggaatgaagtcccgag	839
Chr1	122181	GCACAGCGTGGGCATATTGGCCATGCTGTGGAGAACCGTTTCATTGGACAGAAATCCCGAGG	122240
Chr1	840	tatctacagaaaccccaagagacacatctcttaccatgtctcatcttagaatcgaagacct	899
Chr1	122241	TATCTACCGGGGCCACAGAGGCACATCTCTTGTCTACCTCATTTAGACATGAGAGCCTT	122300
Chr1	900	caaccaatgacccggggaagtgtgcgaataatacaaaaagacctgggcttgaattgtgcgagct	959
Chr1	122301	CACCTATGATACCGCGGAAGTGCACAATAATCAAAATAATGCGGGCTTGAATTGTGTGAGCT	122360
Chr1	960	ggtgtataaccggttaacgacct-----agcccttgagtgtgaattgtgtccgcacctgc	1011
Chr1	122361	GGTGTATFACGTGTTTCTGGCTTTCGGGCACACACCCCTGACTGTGATTTGTCCGCCACTGC	122420
Chr1	1012	atcgccaagatcccaagagcagtggaaggaagtgtcagtggtctccgtctcaaggccag	1071
Chr1	122421	ATCGCCGAGTCCCAAGAGAGTGTGAAGGAAAGTGCAGGTGTCTGTCTTAAAGGCCAG	122480
Chr1	1072	gtgtacatctcttcggccgggagttcccactgtctctctataaagaagagctgtgtgagctg	1131
Chr1	122481	GGTGTATCTCTTGGTGGAGTGGCC--CACTGTCTTAAACGAGAGCGGGTGTGACATG	122537
Chr1	1132	aacgtgcgaaggtgtattatgaagccaactgatgcacacgggtctcatcaacaataatccctc	1191
Chr1	122538	AACGTGCAGAGGTGATTAATGACCAATTGATGCCACCGGTTTATCATCAACTCAATTCCCTC	122597
Chr1	1192	aggctgaagaaatcatcatcgtctccagagcaagtgtaactgtgccaaatag	1239
Chr1	122598	AGGCTGAAGAGATATCATCTGCTCTCCAGAGCAAGTCTACTGCAAAATATAG	122645
RESULT	8		
AP003057/c			
LOCUS	AP003057	190665 bp	DNA
DEFINITION	Homio sapiens chromosome 11 clone RP11-382M14 map 11q, WORKING DRAFT		
ACCESSION	AP003057		
VERSION	AP003057.1	GI:11990474	

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens DNA, clone:RP11-382M14.
Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 190665)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Toyokuni, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 190,665 genomic DNA of 11q
Published Only in Database (2000) In press
2 (bases 1 to 190665)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Toyokuni, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (20-DEC-2000) Masahira Hattori, The Institute of Physiological
and Chemical Research (RIKEN), Genomic Sciences Center (GSC):
1-7-22 Suenho-cho, Tsukuba-shi, Ibaraki, 305-0857, Japan
(E-mail: hattori@gscc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>)
Tel: 81-45-503-9111, Fax: 81-45-503-9170
Genome Center

```

Center for RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hpg.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information -----
Center project name: Humdraft11
Center clone name: RP11-382M14
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; Err-0.990329
Consensus quality: 185598 bases at least Q40
Consensus quality: 187598 bases at least Q30
Consensus quality: 188368 bases at least Q20
Insert size: 188765; sum-of-ctrls
Insert coverage: 8.21x in Q20 bases; sum-of-ctrls
-----
NOTE: This is a 'working draft' sequence. It currently consists of
20 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1
31506 contig of 31506 bp in length
1
31607 63890 contig of 32284 bp in length
63991 91527 contig of 27537 bp in length
91628 106175 contig of 14548 bp in length
106276 115509 contig of 9234 bp in length
115610 125111 contig of 9502 bp in length
125212 133211 contig of 8000 bp in length
133312 142573 contig of 9282 bp in length
142674 149569 contig of 6866 bp in length
149670 155256 contig of 5597 bp in length
155367 160596 contig of 5220 bp in length
160697 165212 contig of 4516 bp in length
165313 171081 contig of 5769 bp in length
171182 174820 contig of 3639 bp in length
174921 178093 contig of 3173 bp in length
178184 182424 contig of 4231 bp in length
182525 183733 contig of 1209 bp in length
183834 186908 contig of 3075 bp in length
187009 189277 contig of 2269 bp in length
189378 190665 contig of 1288 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
31506: contig of 31506 bp in length

```

```

* 31507 31606: gap of 100 bp
* 31607 63890: contig of 32284 bp in length
* 63891 63990: gap of 100 bp
* 63991 91527: contig of 27537 bp in length
* 91528 91627: gap of 100 bp
* 91628 106175: contig of 14548 bp in length
* 106176 106275: gap of 100 bp
* 106276 115509: contig of 9234 bp in length
* 115510 115609: gap of 100 bp
* 115610 125111: contig of 9502 bp in length
* 125112 125211: gap of 100 bp
* 125212 133211: contig of 8000 bp in length
* 133212 133311: gap of 100 bp
* 133312 142573: contig of 9262 bp in length
* 142574 142673: gap of 100 bp
* 142674 149569: contig of 6896 bp in length
* 149570 149669: gap of 100 bp
* 149670 155266: contig of 5597 bp in length
* 155267 155366: gap of 100 bp
* 155367 160596: contig of 5230 bp in length
* 160597 160696: gap of 100 bp
* 160697 165212: contig of 4516 bp in length
* 165213 165312: gap of 100 bp
* 165313 171081: contig of 5769 bp in length
* 171082 171181: gap of 100 bp
* 171182 174820: contig of 3639 bp in length
* 174821 174920: gap of 100 bp
* 174921 178093: contig of 3173 bp in length
* 178094 178193: gap of 100 bp
* 178194 182424: contig of 4231 bp in length
* 182425 182524: gap of 100 bp
* 182525 183733: contig of 1209 bp in length
* 183734 183833: gap of 100 bp
* 183834 186908: contig of 3075 bp in length
* 186909 187008: gap of 100 bp
* 187009 189277: contig of 2269 bp in length
* 189278 189377: gap of 100 bp
* 189378 190665: contig of 1288 bp in length.
Location/Qualifiers
FEATURES
source

```

```

1. 190665
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-382M14"
1. 31506
/note="assembly-fragment"
misc_feature
31607..63890
/note="assembly-fragment"
misc_feature
63991..91527
/note="assembly-fragment"
misc_feature
91628..106175
/note="assembly-fragment"
misc_feature
106276..115509
/note="assembly-fragment"
misc_feature
115610..125111
/note="assembly-fragment"
misc_feature
125212..133211
/note="assembly-fragment clone_end:SP6 vector_side:right"
misc_feature
133312..142573
/note="assembly-fragment"
misc_feature
142674..149569
/note="assembly-fragment"
misc_feature
149670..155266
/note="assembly-fragment"
misc_feature
155367..160596
/note="assembly-fragment"
misc_feature
160697..165212
/note="assembly-fragment"
misc_feature
165313..171081
/note="assembly-fragment"
misc_feature
171182..174820
/note="assembly-fragment"

```

```

misc_feature 174921..178093
/note="assembly-fragment"
misc_feature 178194..182424
/note="assembly-fragment"
misc_feature 182525..183733
/note="assembly-fragment"
misc_feature 183834..186908
/note="assembly-fragment"
misc_feature 187009..189277
/note="assembly-fragment"
misc_feature 189378..190665
/note="assembly-fragment"
BASE COUNT 59500 a 36353 c 36659 g 56251 t 1902 others
ORIGIN
Query Match 85.4%; Score 1058; DB 2; Length 190665;
Best Local Similarity 92.2%; Pred. No. 6,1e-210;
Matches 1151; Conservative 0; Mismatches 85; Indels 12; Gaps 3;
QY 1 atgtccagcaagagctccgtgtcttgccctacagctggcgcctggacacctcgtgcatc 60
DB 51242 ATGTCCAGCAAGAGGCTCTATAGTTCTGGCTCTCACTGGCAGCTTGACACTTCTGCATC 51183
QY 61 ctctgtgtgctgaaggaacagctatgactgctatctgccaacttgccaacttgccaag 120
DB 51182 CTCGTGTGCTGAAGGAACCAAGCATATGCTATGCTACCTGGCCAGCATGTGGCCAG 51123
QY 121 aaggaagacttcgaaggaacagcaaggaagcactggaagcttggggccaagaagtgttc 180
DB 51122 AAGGAAGACTTCCAGGAACCCAGGAAGGACGCAAGACCTGGGGCCCAAAAGATGTTA 51063
QY 181 attggagagtcagcaaggaagatttgagagagttcattccttgccggcagcatcagtc 240
DB 51062 ATTGAGAGTCTAAAGGAGAGTTTGTGAGAGTCTGCTGCGCCATGCACTGCACG 51003
QY 241 gactgtatgaagagcactcctcctggcactcctcttgccaagcccttgatcgccgc 300
DB 51002 ACACGTATGAGGACCGGCTACTCTGGGAACCTCTGTCGCGAGCCCTGCATCCAC 50943
QY 301 aaacaagtggaaatcgcccaagcgggaaggggccaagtgtgtcccaagcggccaagga 360
DB 50942 AAACAAGTGGAGTGCAGCCAGCGGAGGGGGCCAATATGTGTCCAGGCAACGCGGA 50883
QY 361 aaggggaagcactcaagctccggttgagctcaagctcactcaactgagccccc-caataaa 419
DB 50882 AAGGGGAAGTCAAGGTCTGGTTGAGCTCACTGCTACTGCTGCTGCCCCACAGATAA 50823
QY 420 ggtcattgtcccttggaagatgctgaattctacaacgggttcaaggcgcgaatgacct 479
DB 50822 GGTCAATTGCTCCTGAAGATGCCGAATTCACAAACGGTTCCAAAGGSCACAGTGA 50763
QY 480 gatgtgaagtaagcaagcaacacggatctccatcccggtcactcccaagaacccgtgag 539
DB 50762 GACGGAATATGCAAGCAACGCGGATTCGCCACCCGGGTACTCCCAAGAACCGTGA 50703
QY 540 catgtatgaaacccatcagcatcagctcagctcagctggaagcttcctgggaagccccaaga 599
DB 50702 CATGAGCAAGACCTCATGCAATCACTGCAAGCTCGAGCTGGAATCTCGGGAATCC 50643
QY 600 ccaacgcctccagtgctctctacacgaagaccgaagaccgaagcccaagcccaacaccc 659
DB 50642 CCAAGGCGCTCCAGCTCTCTACACGAAGCCAGGACCCGCGCAAAACCCCAACACCC 50583
QY 660 tgaacattctcgagatcgagttcaaaaaaggggtccctgtggaagtgcaccaatgcaaga 719
DB 50582 TGACATTCTCGAGATCGAGTTCAAAAAAGGGGTCCCGCTGAAGGAGCAACGTCGA 50523
QY 720 tggaccacacccacagacctcttgagctctctgaactcagcaagcaagtcgagggcaa 779
DB 50522 TGAAGTGCACCCACGACCTCTTGGAGCTCTTCAATGACTGAAAGCAAGTGCAGGCA 50463
QY 780 gcatgagctgggacgtatgacatcgttgagaacacgcttcatltgaaatgaagtcaggag 839

```

Db 50462 GCACAGCGCTGGGCAATATTCGATCGTGAGACACCGTTTCATGAGCAAGATCCGACG 50403
 Oy 840 tatctaagaagaccagagagacacatctcttaccagtgctatttagacatcgagcctt 899
 Db 50402 TATCTACCGGGCCCGACGAGCACCATTCTTGGCTACGCTATTAGCATCGAGGCGCT 50343
 Oy 900 caccatgacccgggaagagcgcaaatcaacagagccttgagctgaattgctgagct 959
 Db 50342 CACCATGTACCGGAGAGTGCAAAATCAACATGCCGGGCTTGAAATTTGCTGAGCT 50283
 Oy 960 ggtgtataccgggttaagcgtct-----agccctgagtgtaattgttcgcgcactgc 1011
 Db 50282 GGTGTATCTAGTGGTTCTGCTGCTTCGCGACACCGCTGAGTGTGATTTGTCGCCACTGCG 50223
 Oy 1012 atcgccagagtcaccagagagcgagtggaaggaagtgacgagtgctcgtctcaagggcag 1071
 Db 50222 ATCGCCGAGTCCCGAGAGTGAGTGAGGAAAGTGACAGGTCTGCTGCTTAAAGGCGCAG 50163
 Oy 1072 ggtacacatcctcgccgagagtcctcctctacatgagagcgagtgtagatg 1131
 Db 50162 GTGTATCATCTTGGCTGGAGATGCGC---CACTGTCTTACACGAGAGCGCGTGACGATG 50106
 Oy 1132 aacgtgcaggtgattatagaccactgagtcacacgggttcacatcaatcattcctc 1191
 Db 50105 AACGTGACGAGGTGATTATGACCAATTGATCCACCGGTTTCATCAACATCAATTCCTC 50046
 Oy 1192 aggcgaaggaatcatcatcgtctcagagagagtgctactgccaatag 1239
 Db 50045 AGCGTGAAGGANTATCATCTGCTCCAGAGCAAGTCACTGCAAAATAG 49998

RESULT 9
 AC073169
 LOCUS AC073169 201989 bp DNA HMG 06-MAY-2001
 DEFINITION Homo sapiens chromosome 11 clone RP11-87E1 map 11, WORKING DRAFT
 AC073169
 VERSION AC073169.1 GI:8389521
 KEYWORDS HMG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 201989)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Bouguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campoliano,A., Castelle,A., Choepey,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArliano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardina,M., Ginde,S., Goyette,M., Graham,L.,
 Girard-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPherson,R.,
 Meldrum,T., Meneses,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Teefage-Stamann,N., Theodor,J., Tittrell,A., Travers,M., Trigglio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 201989)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Bouguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campoliano,A., Castelle,A., Choepey,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArliano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardina,M., Ginde,S., Goyette,M., Graham,L.,
 Girard-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPherson,R.,
 Meldrum,T., Meneses,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Teefage-Stamann,N., Theodor,J., Tittrell,A., Travers,M., Trigglio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (09-JUN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

Project Information
 Center project name: 110271
 Center clone name: 87_E1

Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731

Consensus quality: 189608 bases at least Q40
 Consensus quality: 195790 bases at least Q20
 Consensus quality: 198358 bases at least Q20
 Insert size: 205000; agarose-fp

Insert size: 200189; sum-of-contigs
 Quality coverage: 4.2 in Q20 bases; agarose-fp
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 19 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1275: contig of 1275 bp in length
 1276 1375: gap of 100 bp
 1376 2986: contig of 1611 bp in length
 2987 3086: gap of 100 bp
 3087 4668: contig of 1582 bp in length
 4669 4768: gap of 100 bp
 4769 7645: contig of 2877 bp in length
 7646 7745: gap of 100 bp
 7746 12379: contig of 4634 bp in length
 12380 12479: gap of 100 bp
 12480 16061: contig of 3582 bp in length
 16062 16161: gap of 100 bp
 16162 21161: contig of 5000 bp in length
 21162 21261: gap of 100 bp
 21262 26219: contig of 4958 bp in length
 26220 26319: gap of 100 bp
 26320 34818: contig of 8499 bp in length
 34819 34918: gap of 100 bp
 34919 40947: contig of 6029 bp in length
 40948 41047: gap of 100 bp
 41048 46162: contig of 5115 bp in length
 46163 46262: gap of 100 bp
 46263 56587: contig of 10325 bp in length
 56588 56687: gap of 100 bp
 56688 66612: contig of 9925 bp in length
 66613 66712: gap of 100 bp
 66713 78573: contig of 11861 bp in length
 78574 78673: gap of 100 bp
 78674 92045: contig of 13372 bp in length
 92046 92145: gap of 100 bp
 92146 109655: contig of 17510 bp in length
 109656 109755: gap of 100 bp
 109756 135749: contig of 25994 bp in length
 135750 135849: gap of 100 bp
 135850 164544: contig of 28695 bp in length
 164545 164644: gap of 100 bp
 164645 201989: contig of 37345 bp in length.

FEATURES

1. 201989
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"

```

/misc_feature /map="11"
/misc_feature /clone="RP11-87E1"
/misc_feature /clone.lib="RRC1-11 Human Male BAC"
/misc_feature 1..1275
/misc_feature /note="assembly_fragment"
/misc_feature 1376..2986
/misc_feature /note="assembly_fragment"
/misc_feature 3087..4668
/misc_feature /note="assembly_fragment"
/misc_feature 4769..7645
/misc_feature /note="assembly_fragment"
/misc_feature 7746..12379
/misc_feature /note="assembly_fragment"
/misc_feature 12480..16061
/misc_feature /note="assembly_fragment"
/misc_feature 16162..21161
/misc_feature /note="assembly_fragment"
/misc_feature 21262..26219
/misc_feature /note="assembly_fragment"
/misc_feature 26320..34818
/misc_feature /note="assembly_fragment"
/misc_feature 34919..40947
/misc_feature /note="assembly_fragment"
/misc_feature clone_end:r7
/misc_feature vector_side:right"
/misc_feature 41048..46162
/misc_feature /note="assembly_fragment"
/misc_feature clone_end:SP6
/misc_feature vector_side:left"
/misc_feature 46263..56587
/misc_feature /note="assembly_fragment"
/misc_feature 56688..66612
/misc_feature /note="assembly_fragment"
/misc_feature 66713..78573
/misc_feature /note="assembly_fragment"
/misc_feature 78674..92045
/misc_feature /note="assembly_fragment"
/misc_feature 92146..109655
/misc_feature /note="assembly_fragment"
/misc_feature 109756..135749
/misc_feature /note="assembly_fragment"
/misc_feature 135850..164544
/misc_feature /note="assembly_fragment"
/misc_feature 164645..201989
/misc_feature /note="assembly_fragment"
BASE COUNT 65404 a 37765 c 36760 g 60255 t 1805 others
ORIGIN

```

Query Match	85.4%	Score 1058	DB 2	Length 201989
Best Local Similarity	92.2%	Pred. No. 6e-210		
Matches 1151	Conservative	0	Mismatches 85	Indels 12
				Gaps 3
Qy 1	atgtccagcaaaagtcgcgtggtctctggccctacgtgagcgccttgacacctcgtgcac	60		
Db 19606	ATGTCCAGCAAAAGGCTCTATAGTTTGTGGCCTCAAGTGGCAGCCTTGACACCTTTCATC	196665		
Qy 61	ctcgtgttgctlaagaaacaaagcgtactaacgtcattgtccatctcttggccaacattggccag	120		
Db 19666	CTCGTGTGGCTCAAGGACCAACAGCTATGACGTCATTGCTACTCTGGCCAGCATTTGGCCAG	196725		
Qy 121	aaggaagacttcgaggaagccagaagaagcactgaagcttggggccaaaaagtgctc	180		
Db 196726	AAGGAAGACTTTCAGGAAGGCCAGGAAGAAGCCACTGAAACCTGGGGGCCAAAAAGATGTTA	196785		
Qy 181	attggagatgtcaagcaaggagcttcttggagaggttcattcttggccggtccatccagttccagc	240		
Db 196786	ATTGGAGATGTCACACAGGAGTTTGTGGAGGATTCATCTGGCCGCGCCATGAGATCCACG	196845		
Qy 241	gcacgtataggaagccgctcactctccgtcggacactctcttgcacagccctgaatgcgcgcgc	300		
Db 196846	ACACTGTATAGGAGCCGCTACTCCTCTGGGAACTCTTCTGCCAGGCGCTGCATCAACCCAC	196905		

OY	301	aaacaatgagaaatcgcgcacgagcgagggagggccaaagtatgtgtccacgagcgcccaagga	360
Db	196906	AAACAGTGGCAAGTCGCCCAACGGGAGGGGGCCAAATATGTGTCCACGACACCAAGGGA	196965
OY	361	aaggggaacgaatcagatgcctcggttttagctcagctgtactcaacttggccccc-cagataaa	419
Db	196966	AAGGGGAATGATTCAGTGCTGTGTGTGAGCTCACCTGCTACTGCTGGGCCCCACAGATATAA	197025
OY	420	ggtcaattgtccctctgagagatgtcgtgaattctacaacggttcaagggccgaatgaact	479
Db	197026	GGTCATTGTCTCCCGAAGTATGGCCCGGAATTCACAAACGGGTTCAAGGCCACAGTACCT	197085
OY	480	gatgagagacgcaagaacacacggagattcccatcccggttaacccccaagaaccgctggag	539
Db	197086	GACGGAATATCCAAAGCAACCCGGGATTTCCACCCCGTTACTTCCAGAACCCCTGGAG	197145
OY	540	catgatltagaacctcatgtacatcagactcagactcagagctgtgaattccttggagaaccccaaga	599
Db	197146	CATGAGACGAAACCTCATGTGCACATCAGCTGGAGGCTGGAATCCTGGAGAAATCCGAGAA	197205
OY	600	ccaagcgctccaggtctctctacacgaagaccagaccgaacccaagcccccaaccccc	659
Db	197206	CCAAAGCCGCTTCACAGTCTCTACAGAGAACCCAGAGACCCGGCCAAAAACCCCAACACCCC	197265
OY	660	tgaattctcgaagatcgaattcgaatcaaaaaggggtccctgtaaaggtgaacgaagtcgaaga	719
Db	197266	TGACATTTCTCGAAGATCGAGTCAAAAAAGGGGTTCCCGTAAAGGTGACCAACTCAAGGA	197325
OY	720	tggcaccacccacacagacctccttggagctctctcatgtacctgaacgaagtctgcggagca	779
Db	197326	TAGTGGCACCCACAGACTCTTGGAGCTTTCATGTACTGACTGACAGAAATCGCGGGCAA	197385
OY	780	gcatgtcgctggcgccgatttgaatcgtgtggaagacgcttcaattggaaatgaagtcccgagg	839
Db	197386	GCACAGCGCTGGGCAATATTGGCCATCGTGGAGAACCGTTTCATTGGAAAGCAAGTCCCGAGG	197445
OY	840	tatctacgaagccccagagagacacatccctttacacgtcattttaagaatgaagacct	899
Db	197446	TATCTACCGGGGCCCGACGAGGACCATCTCTTGGCTAGCGCTATTATGACATCGAGGCTTT	197505
OY	900	caccaatgacccgggaagtgtgcgcaaatccaatccaagagcccttggcttgaatttctgaact	959
Db	197506	CACCATGTACCGGGGAGTGCACAAATCAAAACATGGCCGGGGCTGTGAATTTGCTGAGCT	197565
OY	960	ggtgtataccggttaacgacct-----agccctgaagtgtgaatttgtgcgcacatgc	1011
Db	197566	GGTGTATATCTGGTTTCTGTTCTGGCTTCCGGACACAGCCCTGATGATTTGTGCCCACTGC	197625
OY	1012	atcgccaaagtcccgagagcgagtggaaggaagaatgtgaagttgtccgtccccaagggccag	1071
Db	197626	ATTCGCGGAGTCCCAAGAGATGAGTGAAGGGAATGTGAGGCTGTCTTTAAAGGCGAG	197685
OY	1072	gtgtacatccctcgcgcgaggaatccccaactgtctctctacaaatggaagctgtgtgaacatg	1131
Db	197686	GGTATACATCTTGTGGTGGGAATGGCC---CACTGTCTACCAAGAGAGCCGGTGAAGATG	197742
OY	1132	aacgtgcaggtlgtatltagccaactgtatgtccacggttcatcaacatcaatccttc	1191
Db	197743	AACGTGACGGGTATTAAGACCAATGTGATGCCACCGGTTTCATCAACATCAATTCCTTC	197802
OY	1192	aggtcgaaggaatatactcgtctctccagagcaaggttaactgtccaaatgt	1239
Db	197803	AGGCTGAAGGAATATCATCTGTCTCCAGACGAAGGTCTACTGCAAAATG	197850

RESULT	10		
AL353717/c			
LOCUS			
DEFINITION	AL353717 15019 bp DNA	HTG	05-AUG-2001
ACCESSION	Homo sapiens chromosome 9 clone RP11-562M8, 3 unordered pieces.	***	SEQUENCING IN
VERSION	AL353717		
KEYWORDS	AL353717.9 GI:15131977		
	HTG; HTS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP		

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 155019)
 AUTHORS Sehra, H.
 TITLE Direct Submission
 JOURNAL Submitted (04-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Aug 9, 2001 this sequence version replaced gi:15131481.
 COMMENT Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 Project Information
 Center project name: BA562M8
 Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: M13; M77815; 0% of reads
 Sequencing vector: plasmid; 108752; 99% of reads
 Chemistry: Dye-terminator ET-amersham; 0% of reads
 Chemistry: Dye-terminator Big Dye; 99% of reads
 Chemistry: Dye-terminator amersham; 0% of reads
 Consensus quality: 154190 bases at least Q40
 Consensus quality: 154545 bases at least Q30
 Consensus quality: 154709 bases at least Q20
 Insert size: 154819; sum-of-contigs
 Insert size: 165328; 7.3% error; agarose-fp
 Quality coverage: 8.29x in Q20 bases; sum-of-contigs quality
 coverage: 7.89x in Q20 bases; agarose-fp
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 97615: contig of 97615 bp in length
 * 97616 97715: gap of 100 bp
 * 97716 117003: contig of 19288 bp in length
 * 117004 117103: gap of 100 bp
 * 117104 155019: contig of 37916 bp in length.
 FEATURES
 Location/Qualifiers
 1..155019
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone_id="RP11-562M8"
 /clone_lib="RP11-11.2"
 misc_feature
 1..97615
 /note="assembly_fragment:03117
 fragment_chain:1"
 97716..117003
 /note="assembly_fragment:01189
 fragment_chain:1"
 117104..155019
 /note="assembly_fragment:02543
 fragment_chain:1"
 BASE COUNT 42308 a 31372 c 33444 g 47694 t 201 others
 ORIGIN
 Query Match 85.4%; Score 1057.8; DB 2; Length 155019;
 Best Local Similarity 92.5%; Pred. No. 6.9e-210;
 Matches 1146; Conservative 0; Mismatches 87; Indels 6; Gaps 3;
 Db 65396 ATGTCACCAAGGCTCAGTGTCTGCGCTACAGTGCACGCCGACACCTCTTGATC 65337

QY 61 ctctgtgtcgaaggaagaagcctatgacatcattccatctgtgccaacattgcccag 120
 |||||
 Db 65336 CTCGTGTGGCTGGAAGGAACAAGCTATGACATCTATTCCTCACTGGCCAACTGGCCAG 65277
 |||||
 QY 121 aaggaagactctgaaggaagccgaagaaaggaagcactgaagcttggggccaaaagtgttc 180
 |||||
 Db 65276 AAGGAAGACTCTCGAAGAGCCCTGGAAAGAGGACTGAAAGCTGGGACCAAAAAGTGTTC 65217
 |||||
 QY 181 attgaagatctcgaaggaaggttttgggaagattatctcggccgacatccagtcacg 240
 |||||
 Db 65216 ATTGAGATGCGACAGGAGGAGTTGTGGAGAGTTCATCTGGCCGCGCATCCAGTCCAC 65157
 |||||
 QY 241 gcaatgatagaagcagcctacccctcgtgacccctcttcgagccctgacatccgcgcg 300
 |||||
 Db 65156 GCACTGATATGAGAACCCCTACCTCTCTGGGACACTCTGTGGCAGGCCGCGCATCCGCGG 65097
 |||||
 QY 301 aaacaagtgaaatcgccagcggaaggggccaaglatgtgtccacgycgcacagga 360
 |||||
 Db 65096 AAACAAGTGAATAATCGCCCAAGAGAGAGGCCAAGATGTGTTCCACGCGACCGGGA 65037
 |||||
 QY 361 aaggggaagcatgaagtcgggttttggatgaagctgactcactgccccccagataag 420
 |||||
 Db 65036 AAGGGAATGATCAGTCCGCTTTGAGCTCCTGCTATTGCTGGGCCCCCGAGATAAAG 64977
 |||||
 QY 421 gcatctccctcgggaagatgctgaattctacacacggttcaaggccgacatgacctg 480
 |||||
 Db 64976 GTCCACCGCTCTCTGAGAGATGCCGAGTTTACAAACGGTTTCAAGGCTGCAATGACTCG 64917
 |||||
 QY 481 atgagtagcaagaagcaacacggaattccatcccggtcactcccaagaaccggtgagc 540
 |||||
 Db 64916 ATGAAATAGCAAAAGCAACATCGGGTTCCATTCACGTACCTCCAGAGCCGCTGAAC 64857
 |||||
 QY 541 atgagtgaagactcctatgacatcagctgaagagctggaatccttggagaaccccaagac 600
 |||||
 Db 64856 ATGACGAGAACTCATGCAATCAGTACGTAGAGGCTGGAATCTG--GAACCCCAAGAAC 64799
 |||||
 QY 601 caagcgctccagatctctacacgaagacccaagaccgaagccccaacacccct 660
 |||||
 Db 64798 CAACTGCTCTCAGTCTCTCATGAAAGACCCAGGACCCCAAGCCCTCAACACCCCT 64739
 |||||
 QY 661 gacattcagagatcgaatcaaaaaaggggtccctgtgaaggtgacacgttcaagat 720
 |||||
 Db 64738 GACATTCCTCAGATCAAGTTCAAAAAAGGGGTCCTCGTGAAGGTGACCAAGTCAGAGAT 64679
 |||||
 QY 721 ggcacaccccaacgaagactctcttggagctcttactgaactgaagaaatcgccggaag 780
 |||||
 Db 64678 GGCACACCCCAACGAGACCTCTTGAGACTTTTGTGTACTGTAAGATTGACAGCAAG 64619
 |||||
 QY 781 catgacgtggcgctattgacatcgtggaagacccgttcaattggaatggaatcccgagct 840
 |||||
 Db 64618 CACGGCTGTGGCCGTATGTGACATCGTGGAGAACCCGCTTATTTGAAATGAACTCTGAGGT 64559
 |||||
 QY 841 atctaagagaccccaagcagacacatccttaacatgcatcattgatacgaagccctc 900
 |||||
 Db 64558 ATCTGGAGGCGCCACAGAGGACCATCTTCAACACATCATTTAGATCAAGGCGCTTC 64499
 |||||
 QY 901 accatggaccgggaagatgacaaaatcaacaaagcctggtgttgaatttctgaagctg 960
 |||||
 Db 64498 ACCATGTACCGAAGATGCGCAAAATCAAAAGGCTGAGCTGAAATTTCTGAGCTG 64439
 |||||
 QY 961 gtgtataccggtttaaagccttaagccttgaatgtgaattgttcgcaatgacatccgaag 1020
 |||||
 Db 64438 GTGTATACCGGTTTCTGGCATACCCCTGATGTGATTTGTCTCGGCACATGTAATGCAAG 64379
 |||||
 QY 1021 tcccagagcagagtgaaaggaagatgacagtgcttcgctccaagggccagtgatcaltc 1080
 |||||
 Db 64378 TCCACAGGAATGAATGGAAGGAAGTGCAGGTCTGTCTCAAGGCGCAGGTGATATTC 64319
 |||||
 QY 1081 ctggccgagagtcacccactgtctctcaaatgaggaactgtgtgacatgaagctgacg 1140
 |||||
 Db 64318 CTCGGTGGGAGTCCCACTGCTCTCTTACAAAGAGAGAGACTGTGTGACATGAAGCTGAG 64259
 |||||

[illegible]

```

REFERENCE TITLE Direct Submission
AUTHORS Submitted (01-JUN-2001) Human Genome Sequencing Center, Department
JOURNAL Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 159687)
Morley,K.C.
Direct Submission
Submitted (05-JUN-2001) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 1, 2001 this sequence version replaced gi:14190590.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STS are identified using EPCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE:sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
html.

QUALSTAT-REPORT-----
----- Summary Statistics -----
Contig length: 160224
Phrap values in estimate: 159430
Average error rate (BCM-Phrap estimate): 1.12363e-05
Fraction of Phrap values less than 40 : 0.00137364
Number of consensus changing edits: 0
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original+Context Edited+Context
-----
----- Distribution of Quality < 40 Bases -----
2001
1801
1601
1401
#
bases 1201
1001
801
*
*

```

```

601
401
201
01
5 10 15 20 25 30 35 40
Phrap Value Range

```

Version: 1.01 qxf0.

FEATURES

Source

```

Location/Qualifiers
1. 159687
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-267D19"
11. 542
/rpl_family="L1MEC"
34. 137
/standard_name="139279"
593. 700
/rpl_family="L1MEC"
725. 1027
/rpl_family="AlusX"
1056. 1401
/rpl_family="L1MEC"
1419. 1974
/rpl_family="L1ME1"
1985. 2067
/rpl_family="MLT1-INTERNAL"
2141. 2246
/rpl_family="MLT1-INTERNAL"
2421. 2938
/rpl_family="MLT1D"
3967. 3992
/rpl_family="AT_rich"
4687. 4824
/rpl_family="MIR"
5032. 5098
/rpl_family="AT_rich"
5927. 6111
/rpl_family="L1Pba"
6107. 6578
/rpl_family="L1Pba"
6601. 6953
/rpl_family="L1Pba"
6934. 8357
/rpl_family="PTRS"
8358. 8778
/rpl_family="L1Pba"
8779. 9084
/rpl_family="AlusG"
9085. 9710
/rpl_family="L1Pba"
9881. 9929
/rpl_family="GA-rich"
complement(11501..11569)
/rpl_family="MIR"
complement(11571..12240)
/rpl_family="MER21B"
complement(12242..12407)
/rpl_family="MIR"
complement(12510..12804)
/rpl_family="AlusX"

```

Query Match 85.2%; Score 1055.4; DB 9; Length 159687;

Best Local Similarity 92.4%; Pred. No. 2.2e-209;

Matches 1145; Conservative 0; Mismatches 86; Indels 8; Gaps 3;

```

OY 1 atgtccagaagctcgtgtgtctgtgctcaagtgccgctcggaacacctgtgcatc 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55610 ATGTCCAGAAAGGCTGTGTGTCTGTGCTCAAGTGTGTGTGACACCTCCTGCATC 55551

```

```

OY 61 ctctgtgagctgaagaaagcctatgacgtcctctatctctggtccaaattgcccag 120
    ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55550 CTCTGTACTGTGAAGAACAGGCTATGATGTCTTCTTACCTGGCCACATTAAGCCAG 55491

OY 121 aaggaagactctgaggaagcccaagaaagaaagcactgaagcttggggccaaaagtgttc 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55490 AAGGAGATTTCAGGAAACAGAGAGAA -GCATTGAAGCTTGGGGCCAAAAGTGTTTC 55432

OY 181 atgagatgtcagcaagaggttctgtgaggaagttcaatctgcccgcacatccagtcagc 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55431 ATTGAGATGTACACAGAGGAGTTGTGTGAGAGATTCAATCTGGCCAGCATTCAGTTACG 55372

OY 241 gcactgatatgagaccgctacccctctctggtgacccctcttgcagccctcatccggccg 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55371 GCATTCGATGAGACCCCTTACTCTGTGGCACTTCTCTACACAGGCCCTGCTCATGCTGCG 55312

OY 301 aaaaacgtggaatcgcgcagcggaaggggccaagtatgtgtccacgycgcacagga 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55311 AAACAGTGGGAATTGGCCACCGGAGGGGCCAAGATATGTGTCCATGGCCTCAGCGGA 55252

OY 361 aagggaagatcagatcaggtcgggtttagctcagctgactactactggtcccccagataag 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55251 AAGGGAGCGCATCAGTCTCCAGTTTGAAGCTCACTGCTACTCGCTGCTCCAGATAAG 55192

OY 421 gtcatgtccctggaagatgacctgaattctacaaccggttcaaggccgcaatgacctg 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55191 GTCAATGCTCCCTGGAGCATGCTCTGAGTTTACACCGGTGTAAGGGCCCAACGACGCTG 55132

OY 481 atgagatcagcaagcaacacaggaatcccatcccggtacatcccaagaaccgttgagc 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55131 ATGGAATATCGCAAAAGCAGCAGCGGATTCCTATCTGTCTCTCCCAAGAACCTGTGTAGC 55072

OY 541 atgagatgagaacctatgacatcagctacgagctggaatccctggaagaaaccccaagac 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55071 ATGAAACAAGAACCTCATGACATCAGCTAGCTGAGAGTGAATCCTGGAAGATCCCAAGAC 55012

OY 601 caagcgctcagatctctacagaaagcccaagaccgcaaaagcccaaacacct 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55011 CAAAGCCTCCAGGTCTCTCTACGGCAAGACCCAGGACCC-----GGCCCCCAACACCCCT 54958

OY 661 gacatctcagatcgaatcgaatcaaaaaggggtccctgtggaagtgcacaaagtcgaagat 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54957 GACATCTCGAGATTGAGTT -AAAAAAGGGGTCCCGGTGAAGGTGACTGACATCAAGAT 54899

OY 721 ggcacaccccaagacacctcttggagctctcaatgtaacctgaagaagtgcgggaag 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54898 GGCCTCCACCTACACAGACCTCTTGAGCTTCTCATGTACTGAAACGAAGTCCGGGGCAAA 54839

OY 781 catggcgtggccgtatctgacatcgctggaagaaacgcttcatctggaatggaatcccgaggt 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54838 CAGGGCTGGGGCGGTACTGACATCTGTGAGAACCCGCTTATTGGAATGAAGTCTCAGGT 54779

OY 841 atctacgagaccccaagcagacacatccttaccatgctcatattagacatcgagccttc 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54778 ATCTAGAGAGCCCGGAGGACACATCTTTTACACGCTCATATTGACATCAAGGCTTTC 54719

OY 901 accatggaccgggaagtgtgcacaaatcaaaagagcgtggtgcttgaatttctgagctg 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54718 ACCATGAGCTGGGAAGTGCACAAATCAAAAGGCGCTGGGCTGGAATTTCTGTGAGCTG 54659

OY 961 gtctataccggttaccgttaccctagcctgtgagttgtgctgcacagcagtcacgcaag 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54658 GTATATACCGGTTTCTGTGGCAGACACTGAGTGTGATTTTCTGCGCAGCTTCGCCAAG 54599

OY 1021 tcccagaagcgaatggaaggaagatgacagtggtccgtccctcaagggccaaagtgtacatc 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54598 TCCAGAGAGCAGTGAAGGAAGGAAGTGTGTCTGTTCAAGAGGCCAAGGTGATCAATC 54539

OY 1081 ctctggcggaggtccccaatgtctctcacaatgagagcgtgtgagcaatgagctgcaag 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54538 CTCTGGCGGTAGTCCCACTGCTCTCTCTACATGAAGAAAGCTGTGAGCTGAAACATGCGAG 54479

```


OY 1141 ggtatgttgagccacgacgagcgggtatcatcaatcattccctcgaagcctgaag 1200
|||||
Db 54478 GGTGATTATGAGCCATATGCTGCTTCATCAACATCACTCCCTCAGGCTGAG 54419
|||||
OY 1201 gaattcatcgtctccagagcaagtcactgcaaatag 1239
|||||
Db 54418 GAATATCATCATCTCCAGAGCAAGTCTCTCCCAATATG 54380
|||||
RESULT 12
AC026357/c
LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-85503, WORKING DRAFT
ACCESSION AC026357
VERSION AC026357.22 GI:12044426
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 204371)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alshrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonin,D., Bouck,J.,
Bowe,S., Brieleva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hayes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C.,
Hollins,B., Homs,J.F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Lounsged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,K., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogub,M., Okunolu,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Picken,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuboan,I., Rolfe,M.,
Rulz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Soneike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellier,B., Thomas,N.,
Thomas,S., Usmani,K., Vasequez,L., Vera,V., Villalobos,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williams,A., Wleciyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 204371)
Worley,K.C.
Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 8, 2001 this sequence version replaced gi:12043803.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HASE
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 94% of reads
Chemistry: Dye-terminator Big Dye; 94% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 201671 bases at least Q40
Consensus quality: 205669 bases at least Q30
Consensus quality: 205669 bases at least Q20
Estimated insert size: 201838; sum-of-coverage
Quality coverage: 0x in Q20 bases; agarose-rip estimation
Quality coverage: 9.7x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 8 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 52254: contig of 52254 bp in length
52255 52354: gap of unknown length
52355 101848: contig of 49494 bp in length
101848 101948: gap of unknown length
101948 134385: contig of 32437 bp in length
134385 134485: gap of unknown length
134485 164948: contig of 30463 bp in length
164948 165048: gap of unknown length
165048 191727: contig of 26679 bp in length
191727 191827: gap of unknown length
191827 201957: contig of 10130 bp in length
201957 202057: gap of unknown length
202057 203210: contig of 1153 bp in length
203210 203311: gap of unknown length
203311 204371: contig of 1061 bp in length.

FEATURES
source
1. 204371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-85503"
BASE COUNT 64581 a 39919 c 38928 g 60229 t 714 others
ORIGIN

Query Match 85.2%; Score 1055.4; DB 2; Length 204371;
Best Local Similarity 92.4%; Pred. No. 2.1e-209;
Matches 1145; Conservative 0; Mismatches 86; Indels 8; Gaps 3;

OY 1 atctcgaagaagggccggtgttctggtcctacagtggtggtggaacccctgcatc 60
|||||
Db 158529 ATGTCCAGAAAGGCTCTGTGTTCTGCGCTACAGTGGGCTGAGACCTCTGCAATC 158470
|||||
OY 61 ctctgtgtgctgaaggaagaagcctatgacgtcattgctcctatctgccaacattggccag 120
|||||
Db 158469 CTCCTGTACTGCTGAGGAAACAAAGCTATATGTCATTTGGCTACTGCGCAACATTGCGCAG 158410
|||||
OY 121 aaggaagacttgaggaagcaggaagaagcactgaagccttgaggccaagaagtgctc 180
|||||
Db 158409 AAGGAAGATTTCAGAGAAACCGAAGAA- GCATTGAAGCTTGGGGCCCAAAAAGGTTGTC 158351
|||||
OY 181 attgagagatcagcagagagatttgtagagagattcattcctgctgagcagcagc 240
|||||
Db 158350 ATTGAGAGATGTCAGCAGGAGAGTTTGTGAGAGAGTTCTGCGCCAGCCATCAGTTTCAGC 158291
|||||

Db	33680	GGTCAATGCTCCCTGGAGGATGGCCAAAGTTTCACAAACAAGTTTCAAGAGTCCCAAAATGACT	33739
Qy	480	gatgagtaagcaaaagcaacagcggaattccatcccggtacatcccaagacccgtgag	539
Db	33740	TATGGAAACACAAAGCAACGGGATTTCCATTCCTCACTCCCAAGAACTGTGGAA	33799
Qy	540	catgtatgagaacctcatgacatcagctacagagcttgaaacctgtgaagaccccaagaa	599
Db	33800	CATGAGAGAGAACCTCTATGAGATTCAGCAATGAGCTGTGAATCTTGGAAACCTTAGAA	33859
Qy	600	ccaagcgccctccaggtgtctctaaagcgaagcccaagaccagcaaaagccccaaccccc	659
Db	33860	CCAAGCAATTTCCAGGTCTCTCCAGAAAGCCAGACCCGGCCAAAGCCCCAAGACCC	33919
Qy	660	tgacatctcagagatcgafttcaaaaaagggctccctgtgaagtgatcaacagtcacaaga	719
Db	33920	TGACATTTCTGAGATGAGATTGAATTAAGAGTCCCATGAAAGTGACCAAGTCMAAGA	33979
Qy	720	tggacacacccacacagacactcctttgagctcttcaatgtacacttgaaagagtcgaggaa	779
Db	33980	TGGCACCCACCCACAGACCTCCTTTGAGGCTTTCCTGTACTGAAACGAAGCTGGGGCAA	34039
Qy	780	gcacgtggtggccgattgaacatcgtagaagacgcgtctatlgaaatgaagtcccag	839
Db	34040	GTAACGTTGGCGGATTTGACTCAAGSAGAACCACTTCACTGGAATGAAGTCCAGG	34099
Qy	840	tatctacgaaccccaagcagacacatcccttaccatgtcatltagacatcgaagcct	899
Db	34100	TATCTTAGAAGACCCCGAGACACCATCGTTTACCATGCTCATTTAGACATCAGGCTT	34159
Qy	900	caccatgagccgggaagtgcgaataatacaaaagccttgagcttgaaatttgcgtagct	959
Db	34160	CACCATGACCGGAAAGTATGCAAAATCAAAATGAAGCCTGGGCTTTGAAATTTGCTGAGCT	34219
Qy	960	ggtgtataccggttaccgcccagcccctgaagtgtaatcttttgcgcaactgcatcgcgaa	1019
Db	34220	GGTGTAACGGGTTCTGGCACAGCGCTTGAAGTAAATTTGTGCCACCACTGTCATTGCCAA	34279
Qy	1020	gtcccaagagcgaagtggaaaggaaagtgcgaagtgctcgtctctcaaggccaggtgtacat	1079
Db	34280	GTCOCAGAGATGAGTGGAAAGGAAAGTGCAGGCTGTCCTTTCAAGGGCCGTGTTACAT	34339
Qy	1080	ccctgcgcgggaatctcccaactgtctctctcaaaatgaagaagctgtgtgacatgaacgca	1139
Db	34340	CTTAGCGCAAGACT-CCCTCTGTTTATCTTACACAGSAGAGCTGTGAGCAAGAACGTGCA	34398
Qy	1140	gggtgtatlatgagccaactgatatgcacccggtgtctcaacaatcaatccctcaagctgaa	1199
Db	34399	GGTGTATTATGAGCACAATGATGATGCCACAGTTTCATCAACATCAATTCCTCAGAGCTGA	34458
Qy	1200	ggaatatcatcgtctccagagcaagtgatcactgcgcaaatag	1239
Db	34459	GGATATCATCATCTCCAGAGAGGTCACTGCACCAAAATAG	34498
RESULT 14			
HUMAS3PS			
LOCUS	HUMAS3PS	1868 bp	DNA
DEFINITION	Human argininosuccinate synthetase pseudogene 3.		
ACCESSION	K01846		
VERSION	K01846.1	GI:179058	
KEYWORDS	argininosuccinate synthetase; processed gene; pseudogene; synthetase.		
SOURCE	Human genomic DNA, Charon-4A library of Lawn et al, clone psi-AS-3		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1868)		
AUTHORS	Freitag,S.O., Bock,H.G., Beaudet,A.L. and O'Brien,W.E.		
TITLE	Molecular structures of human argininosuccinate synthetase pseudogenes. Evolutionary and mechanistic implications		
JOURNAL	J. Biol. Chem. 259 (5), 3160-3166 (1984)		

MEDLINE COMMENT		FEATURES	
<p>84135624</p> <p>There is one expressed gene and 14 pseudogenes for argininosuccinate synthetase in the human genome. Three of the pseudogenes (psi-AS-1, -3, -7) were sequenced, in addition to the cDNA (AS). The pseudogenes are 89-93% homologous to the cDNA and lack introns. The authors [1] estimate that psi-AS-1 and psi-AS-3 were created 10-11 million years ago (93% homology), and psi-AS-7 was created approximately 21 million years ago (89% homology). The functional gene contains 13 exons spanning 65 kb on chromosome 9 (9q34-qter locus).</p>		<p>Location/Qualifiers</p> <p>1. .1868</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/map="9q11-q22"</p> <p>1. .1868</p> <p>/gene="ASSP3"</p> <p>1. .1868</p> <p>/gene="ASSP3"</p> <p>/note="G00-119-709"</p> <p>BASE COUNT 464 a 501 c 474 g 429 t</p> <p>ORIGIN 329 bp upstream of a PstI site.</p>	
<p>Query Match</p> <p>Best Local Similarity 84.9%; Score 1052.2; DB 9; Length 1868;</p> <p>Matches 1143; Conservative 0; Mismatches 88; Indels 8; Gaps 3;</p>			
QY	1 atgtcccaagaagctcgtgtgtcttctgtccctaaagtagtcgagcgctcgtgacacctgtgcatc 60		
Db	262 ATGTCCCAAAAAGGCTGT 321		
QY	61 ctctgtgtgtcgttaagaaacaaagctatagctatcgtcatctgctcctatctgtccaaacatltgacag 120		
Db	322 CTCCTCGACAGTGAAGAGCAAGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATG 381		
QY	121 aaggaagaccttcgaaagccaggaagaaggaagcacttgaagcttggagccaaaagtgcttc 180		
Db	382 AAGGAAGATTTTCAMAGCAACAGAGAGAA- GCATTGAAAGCTTGGGGCCAAAAGGCTGTTC 440		
QY	181 attgaagatgtcaagaaggaatttttgaagaagttcatctatctgtgcgcgcacatcaatgcacag 240		
Db	441 ATTGAAGATGTGACAGCAGGAGATTGTGTGAGAGAGTTTCATCTGTGCGACGACATCCAGTTCACG 500		
QY	241 gcaactgtatlgagaagccgtactcctctgtggcaactctcttgcgaagccctgtgactgcgcgc 300		
Db	501 GCACCGCATGTGAGAGATCGCTACTCTCTGGGCACTTCTTCCACAGGCGCCTGCATCGCTGCG 560		
QY	301 aaaaagaatggaatctgcgcacgcgggaaggggccaagtatgtgtcccaagcggccacagga 360		
Db	561 AAAACAAGTGAATAATTTGCCACCGGGAGGGGGCCAAAGTATGTGTCCATGGGCTCACGGGA 620		
QY	361 aagggaaacgtatcagatgtcagtttgaagtcacgtgtactactcaactgacgtcccccagaataag 420		
Db	621 AAGGGGAGCGATACAGGTCGATTTGAGGCTCAACGTGCTACTGCTGTGCTGCCACAGTAAAG 680		
QY	421 gtcaattgtccctcgtgaggaatgctgtgaatttacaacccggttcaaggcgcgaatgacctg 480		
Db	681 GTCATTGTGCTCCCTGAGAGATGCTGAGATTGTACAAACCGGTTCGAAGGGGCAATAATGACTGTG 740		
QY	481 attgaagtacgcaagaacacacgggatttccatcccggtgtcatctcccaagaacccgtgtgacg 540		
Db	741 ATGGAATATGCGAAMAGCACACGAGGATTTCCATCTGTGTGTCATCTCCCAAGAACCTGTGGAGC 800		
QY	541 attgaatgagaacacctcatgtcacatcagctacgaagctgtgaaatcctctgagaagaccacaagac 600		
Db	801 ATGGAACAAGAACCTCATGACATCATGACGACGAGCTGTGAATCTGTGAGAAATCCCAAGANAC 860		
QY	601 caaagcgctccaagttctcttacaagaagaacccaagaacacagcaagaagcccccaaacacct 660		
Db	861 CAACGCGCTCCAGGTCTCTACGCGAAGACCCAGGATCC-----GGCCCCCAACACCCCT 914		
QY	661 gacattctcgagatcgaatttcaaaaagggttcccttctgaaagtgtaaccacgttcaaggat 720		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:42:27 ; Search time 146.6 Seconds

(without alignments)
7245.743 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239

Sequence: 1 atgtccagcaagagctccgt.....gcaaggtacatcccaatag 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 segs, 428662619 residues

1861242

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /N_Geneseq_1101.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1232.6	99.5	1725	AAH72909	Human cervical can
2	1232.6	99.5	2810	AAH34821	Human colon cancer
3	1232.6	99.5	2812	AAH6219	Human prostate can
4	466.8	37.7	969	AAI12589	Probe #2522 for ge
5	466.8	37.7	969	AAI33937	Probe #2623 used t
6	466.8	37.7	969	AAI02494	Probe #2485 used t
7	450.4	36.4	678	AAI21777	Probe #11710 for g
8	450.4	36.4	678	AAI47060	Probe #15746 used t
9	450.4	36.4	678	AAI07462	Probe #7453 used t
10	351.4	28.4	1048	AAI14080	Aspergillus oryzae
11	310.4	25.1	360	AAI28968	Colon tumour relat

12	278.6	22.5	440	21	AAAC98292	Human ovarian carc
13	229.2	18.5	234	21	AAAT0071	Human secreted pro
14	227.2	18.3	270	21	AAAC30247	Human secreted pro
15	222	17.9	222	22	AAI29415	Colon tumour relat
16	217	17.5	217	22	AAI29308	Colon tumour relat
17	204.6	16.5	1653	21	AAAC34397	Arabidopsis thalia
18	166	13.4	175	21	AAZ80429	Human colon cancer
19	166	13.4	175	21	AAZ80726	Human colon cancer
20	158	12.8	269	22	AAH71231	Human cervical can
21	156	12.6	250	22	AAH72308	Human cervical can
22	143.8	11.6	259	22	AAH69430	Human cervical can
23	134.6	10.9	1203	22	AAH66511	C glutamicum codin
24	134.6	10.9	1209	22	AAH66512	C glutamicum codin
25	134.6	10.9	1326	22	AAH71920	Corynebacterium gl
26	134.6	10.9	1326	22	AAH68528	C glutamicum codin
27	114	9.2	127	19	AAAI1829	Human biallelic po
28	96.4	7.8	1100	22	AAH52904	S. epidermidis ope
29	96.4	7.8	3153	22	AAH54077	S. epidermidis gen
30	65.6	5.3	855	18	AAV74785	Staphylococcus aur
31	56.2	4.5	400	18	AAV75150	Staphylococcus aur
32	50	4.0	92934	21	AAAB1473	N. meningitidis pa
33	50	4.0	172325	21	AAAF21613	N. meningitidis (s
34	44.6	3.6	1000	22	AAAF91396	Streptococcus pneu
35	42.6	3.4	987	19	AAV52348	Aspergillus oryzae
36	41.8	3.4	874	21	AAFI2057	Streptococcus pneu
37	40.6	3.3	7542	19	AAV52191	Human sapiens mamma
38	40.4	3.3	2277	19	AAV13836	Human telomerase p
39	40.4	3.3	2277	19	AAV05372	Human secreted pro
40	40.4	3.3	2371	22	AAAF91874	Human PRO4397 nucl
41	40.4	3.3	2684	21	AAAC58117	Human cDNA sequenc
42	40.4	3.3	3052	22	AAH13926	Corynebacterium gl
43	39.4	3.2	588	22	AAH71221	C glutamicum codin
44	39.4	3.2	1479	22	AAH68245	Viral-encoded sema
45	39.4	3.2	4707	20	AAH55560	

ALIGNMENTS

RESULT 1	
AAH72909	
ID	AAH72909 standard; cDNA; 1725 BP.
XX	
AC	AAH72909;
XX	
DT	19-SEP-2001 (first entry)
XX	
DE	Human cervical cancer marker nucleic acid 4183.
XX	
KW	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200142467-A2.
XX	
PD	14-JUN-2001.
XX	
PF	08-DEC-2000; 2000WO-US33312.
XX	
PR	08-DEC-1999; 99US-0169681.
PR	21-DEC-1999; 99US-0171350.
PR	14-MAR-2000; 2000US-0189315.
PR	12-MAY-2000; 2000US-0203791.
PR	09-JUN-2000; 2000US-0210600.
PR	21-JUL-2000; 2000US-0220114.
XX	
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	
PI	Schlegel R, Deeds J, Berger A, Zhao X;
XX	
DR	WPI; 2001-375006/39.
XX	
PT	New isolated nucleic acid for diagnosing and treating cervical cancer

XX Claim 25; SEQ ID No 2522; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 969 BP; 249 A; 255 C; 244 G; 221 T; 0 other:

Query Match 37.7%; Score 466.8; DB 22; Length 969;
 Best Local Similarity 88.0%; Pred. No. 6.9e-114;
 Matches 521; Conservative 0; Mismatches 67; Indels 4; Gaps 1;

QY 1 atgtccagaagaagctccgtgtgtctgtgctacagtggtggcctgtgacacccctgtgac 60
 Db atgtccagaagaagctccgtgtgtctgtgctacagtggtggcctgtgacacccctgtgac 60
 QY 61 ctctgt 120
 Db ctctgt 120
 QY 441 ctgt 500
 Db ctgt 500
 QY 121 aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 180
 Db aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 180
 QY 501 aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 560
 Db aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 560
 QY 181 atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 240
 Db atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 240
 QY 561 atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 620
 Db atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 620
 QY 241 gcaatgtatgagaagcagctacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 300
 Db gcaatgtatgagaagcagctacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 300
 QY 621 gccctgtatgagaagcagctacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 680
 Db gccctgtatgagaagcagctacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 680
 QY 301 aagaagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 360
 Db aagaagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 360
 QY 661 aagaagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 740
 Db aagaagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 740
 QY 361 aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 420
 Db aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 420
 QY 741 aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 796
 Db aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 796
 QY 421 gtcattgtccctgt 480
 Db gtcattgtccctgt 480
 QY 797 gtcattgtccctgt 856
 Db gtcattgtccctgt 856
 QY 481 atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 540
 Db atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 540
 QY 857 atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 916
 Db atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 916
 QY 541 atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 592
 Db atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 592
 QY 917 atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 968
 Db atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 968

RESULT 5
 AAI3937 standard; DNA; 969 BP.

XX AAI3937:

XX 17-OCT-2001 (first entry)

XX Probe #2623 used to measure gene expression in human placenta sample.
 XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.
 OS Homo sapiens.
 PN W0200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 2623; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).

XX The present sequence is one such probe. The probes are useful for

XX producing a microarray for predicting, measuring and displaying gene

XX expression in samples derived from human placenta. The probes are useful

XX for antenatal diagnosis of human genetic disorders.

XX Sequence 969 BP; 249 A; 255 C; 244 G; 221 T; 0 other;

Query Match 37.7%; Score 466.8; DB 22; Length 969;
 Best Local Similarity 88.0%; Pred. No. 6.9e-114;
 Matches 521; Conservative 0; Mismatches 67; Indels 4; Gaps 1;

QY 1 atgtccagaagaagctccgtgtgtctgtgctacagtggtggcctgtgacacccctgtgac 60
 Db atgtccagaagaagctccgtgtgtctgtgctacagtggtggcctgtgacacccctgtgac 60
 QY 381 atgtccagaagaagctccgtgtgtctgtgctacagtggtggcctgtgacacccctgtgac 440
 Db atgtccagaagaagctccgtgtgtctgtgctacagtggtggcctgtgacacccctgtgac 440
 QY 61 ctctgt 120
 Db ctctgt 120
 QY 441 ctgt 500
 Db ctgt 500
 QY 121 aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 180
 Db aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 180
 QY 501 aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 560
 Db aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 560
 QY 181 atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 240
 Db atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 240
 QY 561 atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 620
 Db atgagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 620
 QY 241 gcaatgtatgagaagcagctacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 300
 Db gcaatgtatgagaagcagctacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 300
 QY 621 gccctgtatgagaagcagctacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 680
 Db gccctgtatgagaagcagctacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 680
 QY 301 aagaagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 360
 Db aagaagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 360
 QY 681 aagaagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 740
 Db aagaagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 740
 QY 361 aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 420
 Db aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 420
 QY 741 aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 796
 Db aaggagaagcttcggaagcagaagaagcagctgaagctgtgtgtgtgtgtgtgtgtgt 796
 QY 421 gtcattgtccctgt 480
 Db gtcattgtccctgt 480
 QY 797 gtcattgtccctgt 856
 Db gtcattgtccctgt 856

QY	421	gttttgctccctggagatgcccgaattcaaacccggttcaagggccgaatgacct	480
Dy	525	gtcatgctccctggagatgctccgaattctactacggtgcacatggtctgcaatgactg	584
QY	481	atgagatagcgacaagacacaggaattccatcccggtctcatccccaagaaccgtgagc	540
Dy	585	atggaatacacaaagacacatcgatctccatccccaatctcctcaagaaccgtgatac	644
QY	541	atgagtagaacctcatgacatcagctacgag	574
Dy	645	atggacaggaaccctcatgctatcatcagctacgag	678

QY	1	atgccacgaagaagctccggtggtctctgctactaagtgtagcgcttgacaactctgcatc	60
Db	109	atgtccacgaacaagaatccatggtctctgtgcccacagatgtgtgctctgacacactcttcacg	168
QY	61	ctctgtgtgacgaaggaacaagcgtactgacgtctatctccatctgtggccaactctggccag	120
Db	169	ctgtgtgtgctgtaaggaacaaagacatggtgtcatctgtccactgtgccaacatctggccag	228
QY	121	aaggaagactctgagagaagccaaagaaagacactgaaagcttggggccaanaagttgttc	180
Db	229	aaggaagactctgagagaagcaagcaagaagaagcatgacaagcttggggccaanaagttgttc	288
QY	181	attgagatgttcacgaagaggttctgtgaaagattcatctcgccggccatccagctccacg	240
Db	289	attgagaaagtcacgaagagattctgtgaaagattcatctcgccggccatccagctccacg	348
QY	241	gcactgtatgagacgcgtactcctctgtggacccctcttgcagagccctcatctgcgcgcg	300
Db	349	gcctgtatgagagcgtactcctcctgtggacccctcttgcagagccctcatctgcgcgcgc	408
QY	301	aaacaagatggaataatcgcgccagcgggaaggcgccaagatattgtcccaacggcgccaagga	360
Db	409	aaacaagtggaataattaccagcaggaaggcccaagtaagtatccatgtgcatcatgga	468
QY	361	aagggggaacgatcagctccggttctgagctcagctcgtctactcactggtgcccccagataag	420
Db	469	aaggagaacgatcagctccaggtctgagatctactcgtctc---cacttggccccagattaa	524
QY	421	gtcattgtccctctgagagatgacctgaatttaacaacgggttcaaaaggcgccaatggacctg	480
Db	525	gtcattgtccctctgagagatgccccgaatttactactctgctcaatggcgcgaatgacctg	584
QY	481	atggaagtacgaagaacaacgggattccatcccggtctactcccaagaaccgttgagac	540
Db	585	atggaattacacaagaacaactcgattccatccacgtaactcctaagaaccgttgacctc	644
QY	541	atggaatgagaacctcatgcatcatgaactgaag	574
Db	645	atggaatgagaacctcatgcatcatgaactgaag	678

DR WPI: 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -

PS Claim 88: Page 2701; 3161pp; English.

XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organization of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*
CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
CC all specifically claimed in the present invention.

XX Sequence 1048 BP; 196 A; 364 C; 251 G; 237 T; 0 other;

Query Match 28.4%; Score 351.4; DB 21; Length 1048;

Best Local Similarity 62.7%; Pred. No. 2.5e-83;

Matches 566; Conservative 0; Mismatches 331; Indels 6; Gaps 1;

QY 19 gttgtctgctaccagtagtgagcgcctgacactcgtctgctggtgaggaagaa 78
DB 88 gtctgtcttgcctactctgtcggtctgagatcagctgcttctcgtgtgacgag 147
QY 79 caaggtatgacgtlactgtcctatctggccaacattggccaagaagaagactgagaa 138
DB 148 gaaggtacgaaggtgtgtcttctctcctcgcgaatgctgcgcaaggaagacgtgscgct 207
QY 139 gccaggaagaagcactggaagcttggtggccaanaagtgcttcatgtgagatgctcagag 198
DB 208 gttgagagaagaagccctcaagatcgtgtccaagaagatgtgattgagatctgcgcg 267
QY 199 gaatttgtagagatgcatctgcccagccatcagctcagctgatatgagagcgcg 258
DB 268 gaatttgtagagagctgtcttctcctgtccatcagatgcaatgcatatgagggctcgc 327
QY 259 taccctctggcaactctcttctgccaagccctgcatgcccgaacaagtgaaatcgac 318
DB 328 taccctctggcaacagctgtgctgtccgttatcgcgcgcgcgcagatgctgtcgc 387
QY 319 cagcggaggaaggcgaagtgtgtcccaagcgcgaaggaagggaagcatcagctc 378
DB 388 cagcgtgaaagctgcagagttgttcaagcagcgtgtcaccgcaagggttaacagccagctc 447
QY 379 cggttgtagctgactgactacacgagcccccagataaagtaagcttctcctgagag 438
DB 448 cgtttggaactgtcttctctgcatccagccctcctcaatgaatcctgcctcttgagct 507
QY 439 atgacctgaattacacacggttcaagggccgaatgacatgtagtgaagcaagcaaa 498
DB 508 gatcccaagttctcaagcgtttcgtggtgcaacgactctcctcgacatgctgcgcacag 567
QY 499 cagcggatccatccctcgttactcccaagaacccgttgagcatgtgataagaacctcag 558

DB 568 accggatccctgttactcctccactaaggcccaagccctgtgcatgtgagcccaactccgc 627
QY 559 cactcagctagagagcgtggaatcctgtgagaaccccaagaaccaagcgtccacagctc 618
DB 628 cactcagctagagagcgtggttcttgagaagccccaacacacccctccgctgacatg 687
QY 619 tacagaagaccagagaccagcccaagcccccacacccctgacatctcagatcag 678
DB 688 tgcacatgacgcgcgcctctgtagccctcctacagcctgcgcgacatcaccatcag 747
QY 679 ttcaaaaagggtctcctgtgaggtgacccaagctcaagatgacacccaccacagac 738
DB 748 ttgcagcaggtatccccaactagct-----cgtactcccgagaagacataacagcag 801
QY 739 tcttgagcttctatgtlactgaacgaagctcggcggaagcagatggtgtggccgtat 798
DB 802 tccgttgagcttcttcaagcctcacaagctcgtgtacacccacaggtgtgtccgtat 861
QY 799 gacatcgttgagaacgccttcatgtgaaatgaagctccagaggtatctacagaccacga 858
DB 862 gatattgtcagagaacgccttcatcgttcaagagcgtgtgctgtatgactccctgt 921
QY 859 ggcacatccttaccatgcttattagacatcagcagccttcacacatgagccggaagt 918
DB 922 atgacatcctcgcgcgcctcattcgcactcgaaggtctcgttctgtagcgcaggtc 981
QY 919 cgc 921
DB 982 cgc 984

RESULT 11

AA128968
ID AA128968 standard; cDNA: 360 BP.

XX AA128968;

XX 12-OCT-2001 (first entry)

XX Colon tumor related determined cDNA sequence for contig 31.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;

XX Immunogenic; gene therapy; vaccine; colonic cancer; ss.

XX Homo sapiens.

XX WO200149716-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-US35596.

XX 30-DEC-1999; 99US-0476296.

XX 10-JAN-2000; 2000US-0480321.

XX 15-FEB-2000; 2000US-0504629.

XX 06-MAR-2000; 2000US-0519444.

XX 19-MAY-2000; 2000US-0575251.

XX 29-JUN-2000; 2000US-0609448.

XX 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
DR WPI: 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the
XX prevention, diagnosis and treatment of colonic cancer -
XX Claim 25; Page 273; 472pp; English.

CC The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colorectal cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patients own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.

CC Sequence 360 BP; 84 A; 96 C; 103 G; 73 T; 4 other;

Query Match 25.1%; Score 310.4; DB 22; Length 360;
 Best Local Similarity 95.5%; Pred. No. 1.2e-72;
 Matches 317; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 758 accctaagaatcgccggaagcagatgctggtggtggtatgacgtgtggaacccgt 817
 Db 1 accctgaacgaatcgccggaagcagatgctggtggtatgacgtgtggaacccgt 60
 QY 818 tcaatggaatgaatgccggaatgacgtgacccgagccacccatccttaccatg 877
 Db 61 tcaatggaatgaatgccggaatgacgtgacccgagccacccatccttaccatg 120
 QY 878 ctcaattgacatcgagccttcacatgagccggaagtgcgcaaaatcaacaagcc 937
 Db 121 ctcaattgacatcgagccttcacatgagccggaagtgcgcaaaatcaacaagcc 180
 QY 938 tgggttgaatgctgacgtggtgtatgacgtggttgcggttgcggttgcggttgcggt 997
 Db 181 tgggttgaatgctgacgtggtgtatgacgtggttgcggttgcggttgcggttgcggt 240
 QY 998 ttgtccgcactgcatcgccaagtcccaagtcgagtcgagtggaagggaagtgtcaggtgtccg 1057
 Db 241 ttgtccgcactgcatcgccaagtcccaagtcgagtcgagtggaagggaagtgtcaggtgtccg 300
 QY 1058 tectcaaggccaggtgtacatccctcgccgg 1089
 Db 301 tectcaaggccaggtgtacatccctcgccgg 332

RESULT 12
 AAC98292
 ID AAC98292 standard; cDNA; 440 BP.
 AC AAC98292;
 XX 09-MAR-2001 (first entry)
 DE Human colon cancer antigen nucleotide sequence SEQ ID NO: 302.
 XX
 XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnerrary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; anti-infective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;

KW Infectious disease; cardiovascular disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200055351-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05883.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587534/55.
 DR P-PDB; AAB53535.

PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer

PS Claim 1; Page 738; 2104pp; English.

CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnerrary, nephrotropic, anti-infective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.

XX Sequence 440 BP; 116 A; 126 C; 115 G; 78 T; 5 other;

Query Match 22.5%; Score 278.6; DB 21; Length 440;
 Best Local Similarity 96.8%; Pred. No. 3.2e-64;
 Matches 284; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 846 cgaagcccgagcgagccatccttaccatgctcatattagacatcgagccttaccat 905
 Db 9 cgaagcccgagcgagccatccttaccatgctcatattagacatcgagccttaccat 68
 QY 906 ggaacgggaagtgcgcaaaatcaacaagcctggcctggaatcttgcgtgagtgta 965
 Db 69 ggaacgggaagtgcgcaaaatcaacaagcctggcctggaatcttgcgtgagtgta 128
 QY 966 taccggttaccgcttagccttagtggtgatttgcgcgacatgcatgccaagtcaca 1025
 Db 129 taccggttaccgcttagccttagtggtgatttgcgcgacatgcatgccaagtcaca 188
 QY 1026 ggaagcgtggaagggaagtgcaggtgtcgtccccaaggccaggtgtacatccctcg 1085
 Db 189 ggaagcgtggaagggaagtgcaggtgtcgtccccaaggccaggtgtacatccctcg 248
 QY 1086 ccgggaagtcgccactgtctctctacaatgaggaagtcgtgtagcatgaactgtca 1139
 Db 249 ccgggaagtcgccactgtctctctacaatgaggaagtcgtgtagcatgaactgtca 302

RESULT 13
 AAA70071/c
 ID AAA70071 standard; cDNA; 234 BP.


```
XX AA070071;
AC 07-NOV-2000 (first entry)
DT
DE Human ovarian carcinoma antigen polynucleotide SEQ ID NO:382.
XX
KW Human; ovarian carcinoma; ovarian cancer; therapy; diagnosis;
KW tumour antigen; identification; cytostatic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200036107-A2.
XX
PD 22-JUN-2000.
XX
PF 17-DEC-1999; 99WO-US30270.
XX
PR 17-DEC-1998; 98US-0215681.
PR 17-DEC-1998; 98US-0216003.
PR 23-JUN-1999; 99US-0338933.
PR 24-SEP-1999; 99US-0404879.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, King GE, Algate PA, Frudakis TN;
XX
DR WPI; 2000-431589/37.
XX
PT Immunogenic portion of an ovarian carcinoma protein and the nucleic
PT acid encoding it, useful for the diagnosis, prevention and treatment of
PT cancer, preferably ovarian cancer.
XX
PS Claim 18; Page 194; 299pp; English.
XX
CC The present invention describes an isolated polypeptide comprising an
CC immunogenic portion of an ovarian carcinoma protein (or its variants).
CC Ovarian carcinoma proteins, and polynucleotides encoding them, have
CC cytostatic activity and can be used in gene therapy and vaccines.
CC Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines
CC are useful for the prevention, diagnosis and treatment of cancer.
CC preferably ovarian cancer. AA069691 to AA070077 and AA012552 to AA012557
CC represent human ovarian carcinoma polynucleotides and proteins used in
CC the exemplification of the present invention.
XX
SQ Sequence 234 BP; 47 A; 63 C; 63 G; 61 T; 0 other;

Query Match 18.5%; Score 229.2; DB 21; Length 234;
Best Local Similarity 98.7%; Pred. No. 3e-51;
Matches 231; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 662 acatctcgaatcgattcctcaaaagggtcccttgtaaggtaccacgtcaagatg 721
DB 234 ACATTCTGGAGATGAGATTCAAAAAGGGCTCCCTGTGAAGGTACCAACGTCAGGATG 175
QY 722 gcacacaccacagacctcttgagctctcctacgtacgtgaacgaagtcgcgcgcgcgc 781
DB 174 GCACACCCACACAGACCTCTTGTGAGCTCTTCATGTACGTGAAGAGAGTGGCGGCAAGC 115
QY 782 atggcgtggggtcgtatgacatcggtgagaacgcgtcattggaatgaagtcgcgcgcgc 841
DB 114 ATGGCGTGGGCGCTATTGACATCTGTGAGAAACCGCTTCATTGGAATGAAAGTCCCGAGGTA 55
QY 842 tctcagagacccagcagcagcaccatcttaccatgctcatttagaatgagag 895
DB 54 TCTACGAGACCCCGACGACGACACCTCTTACACAGCGCTCATTTAGACATGAGG 1

RESULT 14
AAC30247
ID AAC30247 standard; cDNA; 270 BP.
XX
```

```
AC AAC30247;
XX 06-OCT-2000 (first entry)
DT
DE Human secreted protein 5' EST, SEQ ID NO: 34322.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 34322; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 270 BP; 59 A; 72 C; 83 G; 54 T; 2 other;

Query Match 18.3%; Score 227.2; DB 21; Length 270;
Best Local Similarity 99.1%; Pred. No. 1.1e-50;
Matches 226; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtccagcaaaagctcgtgtgtctgtgctcctacagtgcgcgcttgaaacctcgtgcatc 60
DB 40 atgtccagcaaaagctcgtgtgtctgtgctcctacagtgcgcgcttgaaacctcgtgcatc 99
QY 61 ctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
DB 100 ctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 159
QY 121 aaggaagacttcgaagaagcaggaagaaggaagcactgaagccttgaggccaaagtgcttc 180
DB 160 aaggaagacttcgaagaagcaggaagaaggaagcactgaagccttgaggccaaagtgcttc 219
QY 181 attgagatgtcagcagcaggaagttgtgtgagagttcatctgtgcgcgcgc 228
DB 220 attgagatgtcagcagcaggaagttgtgtgagagttcatctgtgcgcgcgc 267

RESULT 15
AAI29415/C
ID AAI29415 standard; cDNA; 222 BP.
XX
```

```

XX  AA129415;
AC  12-OCT-2001 (first entry)
XX  Colon tumour related determined cDNA sequence for clone R0097.C02.
DE  Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX  Immunogenic; gene therapy; vaccine; colonic cancer; ss.
OS  Homo sapiens.
XX  W0200149716-A2.
XX  12-JUL-2001.
XX  29-DEC-2000; 2000WO-US35596.
XX  30-DEC-1999; 99US-0476296.
PR  10-JAN-2000; 2000US-0480321.
PR  15-FEB-2000; 2000US-0504629.
PR  06-MAR-2000; 2000US-0519444.
PR  19-MAY-2000; 2000US-0575251.
PR  29-JUN-2000; 2000US-0609448.
PR  28-AUG-2000; 2000US-0649811.
XX  (CORI-) CORIXA CORP.
XX  Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI  King GE, Wang T, Jlang Y;
XX  WPI: 2001-441847/47.
XX  Colon tumor associated proteins and nucleic acids useful for the
PT  prevention, diagnosis and treatment of colonic cancer -
XX  Claim 2; Page 400; 472pp; English.
XX  The present invention describes colon tumour associated proteins (I) and
CC  the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC  (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC  (II) may be used in the prevention, diagnosis and treatment of diseases
CC  associated with inappropriate colon tumour associated protein (TCAP)
CC  expression, such as colonic cancer. For example, (I) and (II) may be
CC  used to treat disorders associated with decreased expression by
CC  rectifying mutations or deletions in a patient's genome that affect the
CC  activity of TCAPs by expressing inactive proteins or to supplement the
CC  patients own production of them. Additionally, (II) may be used to
CC  produce the TCAP proteins, by inserting the nucleic acids into a host
CC  cell culturing the cell to express the protein. (II) and its
CC  complementary sequences may also be used as DNA probes in diagnostic
CC  polymerase chain reaction (PCR) and hybridisation assays to detect and
CC  quantitate the presence of similar nucleic acids in samples, and
CC  therefore which patients may be in need of restorative therapy. (I) may
CC  also be used as antigens in the production of antibodies against TCAPs
CC  and in assays to identify modulators of TCAP expression and activity.
CC  Anti-(I) antibodies and antagonists may also be used to down regulate
CC  TCAP expression and activity. The anti-(I) antibodies may also be used
CC  as diagnostic agents for detecting the presence of TCAPs in samples
CC  (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
CC  and AA24494 to AA24523 represent nucleotide and amino acid sequences
CC  given in the exemplification of the present invention.
XX  Sequence 222 BP; 31 A; 49 C; 76 G; 66 T; 0 other;
SQ

```

Query Match 17.9%; Score 222; DB 22; Length 222;

Best Local Similarity 100.0%; Pred. No. 2.4e-49;

Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY  446 aattctacaacggttcaaggcgcaatgacgtgagtgaagcaagaacagcgga 505
DB  222 AATTCTACAAACGGTTCAAGGCGCGCAATGACCTGATGAGTACCAAGCAACAGCGGA 163

```

```

QY  506 ttcccatcccggtcactctcccaagaacccgtggaagcatgagaaacctcagcacatca 565
DB  162 TTCCCATCCCGGTCACCTCCCAAGAACCCGTTGAGCATGAGAACCTCATGCACATCA 103
QY  566 gctacagagctggaatcctctggaagaccaccaagcgcctccaggctctctacagca 625
DB  102 GCTACAGAGCTGGAACTCTTGAGAACCCCAAGAACCAAGCGCTCTCAGGTCTTACAGCA 43
QY  626 agaccagaagcccaagcccaagcccaacaacccctgacatc 667
DB  42 AGACCCAGAGACCCAGCCAAAGCCCCCAACACCCCTGACATTC 1

```

Search completed: February 12, 2002, 13:12:38
Job time: 1811 sec

[illegible]

```

? APPLICANT: Melanie K. Spriggs Michael R. Comeau,
? APPLICANT: Robert F. Dubose, Richard S. Johnson
? TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
? TITLE OF INVENTION: RECEPTOR DNA AND POLYPEPTIDES
? NUMBER OF SEQUENCES: 10
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Janis C. Henry
? STREET: 51 University St.
? CITY: Seattle
? STATE: WA
? COUNTRY: US
? ZIP: 98101
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/181,706
? FILING DATE: October 28, 1998
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/958,598 (converted to a
? APPLICATION NUMBER: Provisional, see below)
? FILING DATE: October 28, 1997
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: --to be assigned-- (USSN 08/958,598
? APPLICATION NUMBER: conversion to Provisional application)
? FILING DATE: October 26, 1998
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Henry, Janis C
? REGISTRATION NUMBER: 34,347
? REFERENCE/DOCKET NUMBER: 2631-A
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (206)470-4189
? TELEFAX: (206)233-0644
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4707 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..4707
?
?
? US-09-181-706-1

```


Patent No. 6174689
GENERAL INFORMATION:
APPLICANT: Spriggs, Melanie
TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
RECEPTOR DNA AND POLYPEPTIDES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C. Henry
STREET: 51 University St.
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: MS-DOS/Windows 95
SOFTWARE: Word for Windows 95, 7.0a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/458,791
FILING DATE: 10-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/958,598
FILING DATE: 28-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2631
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4707 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4707
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-458-791-1
Query Match 3.2%; Score 39.4; DB 4; Length 4707;
Best Local Similarity 51.4%; Pred. No. 0.091;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 125 aagacttcgaggaagcgaagaagcactgaagcttgggccaagaaaggtgttcattg 184
DB 623 AGGACACGAGGAGGGGCGCAGCTGGCCACGAGAGCTGGGGCGCTTCAGCTGTGGAGG 682
QY 185 agagatcgcagcaggaagttgttgaggaagttcattctgcgcgcacatccagtcagcgac 244
DB 683 GCGCGGCGACGCTGCACCTTGtGAGCCTTTCTCTGGAACGGCAGCATCTACTTCCCT 742
QY 245 tgatagagacgcgtactctcctcttggaacctctcttcgcaagccctgcagcccgca 301
DB 743 ACTACCCCTACACTATACAGAGGGGCGCTGGCCACCGGCTGGCCACATGGCGCGCA 799
RESULT 9
US-09-459-066-1
Sequence 1, Application US/09459066
Patent No. 6187909
GENERAL INFORMATION:
APPLICANT: Spriggs, Melanie
TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
RECEPTOR DNA AND POLYPEPTIDES

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C. Henry
STREET: 51 University St.
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: MS-DOS/Windows 95
SOFTWARE: Word for Windows 95, 7.0a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/459,066
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/958,598
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2631
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4707 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4707
US-09-459-066-1
Query Match 3.2%; Score 39.4; DB 4; Length 4707;
Best Local Similarity 51.4%; Pred. No. 0.091;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 125 aagacttcgaggaagcgaagaagcactgaagcttgggccaagaaaggtgttcattg 184
DB 623 AGGACACGAGGAGGGGCGCAGCTGGCCACGAGAGCTGGGGCGCTTCAGCTGTGGAGG 682
QY 185 agagatcgcagcaggaagttgttgaggaagttcattctgcgcgcacatccagtcagcgac 244
DB 683 GCGCGGCGACGCTGCACCTTGtGAGCCTTTCTCTGGAACGGCAGCATCTACTTCCCT 742
QY 245 tgatagagacgcgtactctcctcttggaacctctcttcgcaagccctgcagcccgca 301
DB 743 ACTACCCCTACACTATACAGAGGGGCGCTGGCCACCGGCTGGCCACATGGCGCGCA 799
RESULT 10
US-09-103-840A-2/C
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24

FILING DATE: 04-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: 40403
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8582
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3468 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "Synthetic DNA"
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 1..3468
 OTHER INFORMATION: /product= "Full-length pure maize
 OTHER INFORMATION: optimized synthetic Bt"
 OTHER INFORMATION: /note= "Disclosed in Figure 3 as syn11.mze"
 US-08-459-448A-2

Query Match 3.0%; Score 37.2; DB 2; Length 3468;
 Best Local Similarity 49.0%; Pred. No. 0.35; Mismatches 103; Indels 0; Gaps 0;

Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
 QY 1000 gtccgcaactgcatcgccaagtcccaaggagcgagtggaagtgcaagtgctcgtc 1059
 DB 2485 GACGGCCACAGCCCGCTGGCAACTGTGAGAGAAAGCCCTGTGGCGCAG 2544
 QY 1060 ctcaaggcgagtggtacatctctgcccggaggtcccaactgtctcttacaatgaagag 1119
 DB 2545 GCCCTGGCCCGGTGAACCGCCGCAAGAAAGTGCCGACAAACCGCGAAGCTGTGAG 2604
 QY 1120 ctggtgagcatgaacgtgcaagggtattatgagccaactgtagccacgggttcatcaac 1179
 DB 2605 TGGGAGACCAACATCTGTACAGAGGCCCAAGAGACGCTGACGCCCTTGTGTGAC 2664
 QY 1180 atcaattccctcaagctgaag 1201
 DB 2665 AGCCAGTAGACGCCCTGCAGG 2686

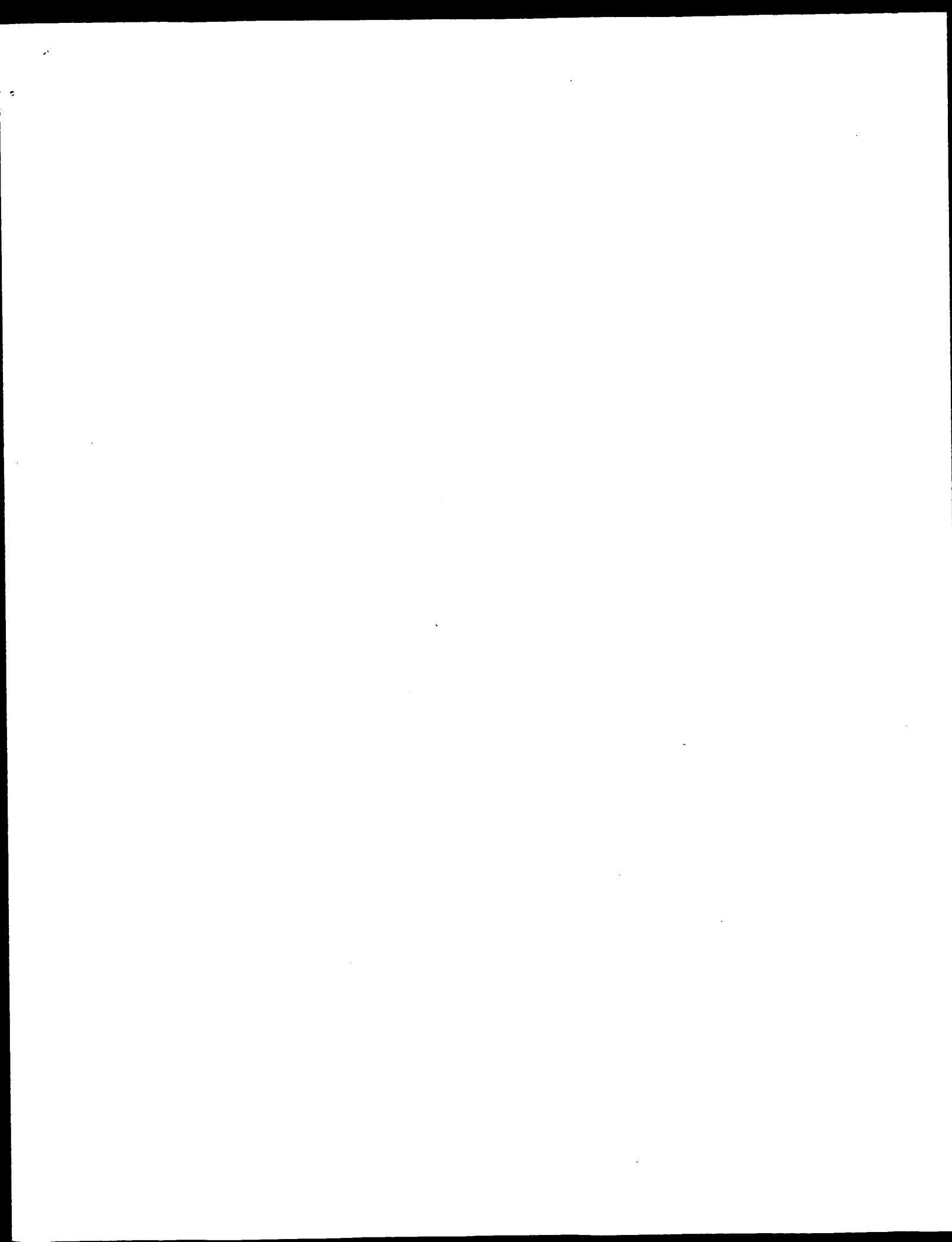
RESULT 15
 US-08-459-448A-4
 Sequence 4, Application US/08459448A
 Patent No. 5859336
 GENERAL INFORMATION:
 APPLICANT: Kozziel, Michael G.
 APPLICANT: Desai, Nallini M.
 APPLICANT: Lewis, Kelly S.
 APPLICANT: Kramer, Vance C.
 APPLICANT: Warren, Gregory W.
 APPLICANT: Ewola, Stephen V.
 APPLICANT: Crossland, Lyle D.
 APPLICANT: Wright, Martha S.
 APPLICANT: Merlin, Ellis J.
 APPLICANT: Launis, Karen L.
 APPLICANT: Rothstein, Steven J.
 APPLICANT: Bowman, Cindy G.
 APPLICANT: Dawson, John L.
 APPLICANT: Dunder, Erik M.
 APPLICANT: Pace, Gary M.
 APPLICANT: Suttie, Janet L.
 TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
 TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: No. 5859336artis Corporation

STREET: Patent & Trademark Dept., 520 White Plains
 STREET: Rd., POB 2005
 CITY: Tarrytown
 STATE: New York
 COUNTRY: USA
 ZIP: 10591-9005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/459,448A
 FILING DATE: 02-JUN-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/951,715
 FILING DATE: 25-SEP-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/772,027
 FILING DATE: 04-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: 40403
 REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIVA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8582
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3468 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "Synthetic DNA"
 HYPOTHETICAL: NO
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 1..3468
 OTHER INFORMATION: /product= "Full length synthetic
 OTHER INFORMATION: maize optimized"
 OTHER INFORMATION: /note= "Disclosed in Figure 3 as syn11.mod. This sequence
 OTHER INFORMATION: identical to flsyn11.fin as disclosed in Figure 1."
 US-08-459-448A-4

Query Match 3.0%; Score 37.2; DB 2; Length 3468;
 Best Local Similarity 49.0%; Pred. No. 0.35;
 Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 1000 gtccgcaactgcatcgccaagtcccaaggagcgagtggaagtgcaagtgctcgtc 1059
 DB 2485 GACGGCCACAGCCCGCTGGCAACTGTGAGAGAAAGCCCTGTGGCGCAG 2544
 QY 1060 ctcaaggcgagtggtacatctctgcccggaggtcccaactgtctcttacaatgaagag 1119
 DB 2545 GCCCTGGCCCGGTGAACCGCCGCAAGAAAGTGCCGACAAACCGCGAAGCTGTGAG 2604
 QY 1120 ctggtgagcatgaacgtgcaagggtattatgagccaactgtagccacgggttcatcaac 1179
 DB 2605 TGGGAGACCAACATCTGTACAGAGGCCCAAGAGACGCTGACGCCCTTGTGTGAC 2664
 QY 1180 atcaattccctcaagctgaag 1201
 DB 2665 AGCCAGTAGACGCCCTGCAGG 2686

Search completed: February 12, 2002, 14:41:32
 Job time: 6940 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:41:27 ; Search time 1390.72 Seconds
(without alignments)
9573.472 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239
Sequence: 1 atgtccagcaagaagctcgt.....gcaaggtcacctgcgaatag 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: qd_estl:*
11: qd_est2:*
12: qd_htc:*
13: qd_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vtl:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	817.8	66.0	878	11	BI256461 602974488
2	809.8	65.4	878	11	BI258834 602869646
3	803.2	64.8	838	11	BI256559 602976921
4	783	63.2	875	11	BI257913 602970844
5	780.6	63.0	815	11	BI260468 602969267
6	780.2	63.0	980	10	BE379161 601238137
7	772.4	62.3	905	11	BE379161 601238137
8	768.8	62.1	918	11	BI259920 602968861
9	758.4	61.2	952	10	BE378581 601236939
10	753.2	60.8	981	11	BI256501 602870143
11	749.2	60.5	1028	11	BE206265 601869276
12	748.6	60.4	898	11	BI253774 602976122

13	745.4	60.2	851	11	BI258628 602969566
14	741	59.8	894	11	BI258439 602872421
15	738.8	59.6	851	11	BI255104 602975950
16	734.4	59.3	901	11	BI254017 602975178
17	733.4	59.2	822	11	BI258153 602971104
18	727.6	58.7	904	11	BE315954 601895904
19	726.6	58.6	1100	10	BE783168 601470778
20	725.6	58.6	900	11	BI253130 602873215
21	719.4	58.1	844	11	BI253130 602873215
22	717.6	57.9	766	10	BI334157 602897591
23	717.4	57.9	916	11	BE899161 601681957
24	715.8	57.8	1012	11	BE437839 602489758
25	715.4	57.7	843	10	BE614762 601281336
26	715	57.7	890	11	BI254203 602974855
27	707.6	57.1	1039	11	BE403394 602819070
28	705.2	56.9	906	11	BI334814 602989331
29	696.6	56.2	836	10	BE745306 601574086
30	696.6	56.2	1094	10	BE745308 601574086
31	696.2	56.2	862	11	BE394443 602456967
32	692.8	55.9	885	10	BE379436 601237366
33	692.2	55.9	1023	11	BE797765 601586321
34	692	55.9	731	10	BE745841 601580056
35	692	55.9	884	10	BE742766 601574370
36	690.8	55.8	820	11	BE400766 602464167
37	687.4	55.5	879	11	BE489990 602518529
38	686	55.4	733	11	BE741651 602631305
39	684.4	55.2	814	10	AU132912 602631305
40	682.2	55.1	763	11	BE792319 601585528
41	680.2	54.9	780	10	BE543488 601070608
42	680	54.9	820	10	BE543057 601068986
43	679.4	54.8	1013	10	BE378897 601237474
44	675.2	54.5	827	10	AU143326 60143326
45	675	54.5	971	10	BE615027 601281264

ALIGNMENTS

RESULT 1
BI256461 878 bp mRNA EST 17-JUL-2001
LOCUS 602974488F1 NIH_MGC.12 Homo sapiens cDNA clone IMAGE:5113965 5',
DEFINITION mRNA sequence.
ACCESSION BI256461
VERSION BI256461.1 GI:14810890
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMW at:
http://image.llnl.gov
Plate: LLM11276 row: 0 column: 22
High quality sequence stop: 835.
Location/Qualifiers
1. 878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5113965"
/clone_lib="NIH_MGC.12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"

FEATURES

source

/note="Organ: cervix; Vector: PCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 219 a 259 c 246 g 154 t

ORIGIN

Query Match 66.0%; Score 817.8; DB 11; Length 878;
Best Local Similarity 98.8%; Pred. No. 4,9e-185;
Matches 845; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

1 atgcccgaagaagctcgggtgtctgtctgacagtcgagccgacccctgcatc 60
|||||
25 ATGTCCAGCAAGCTCGTGTGTCTGTGCTACAGTGGGCGCTGGACCTCTGATC 84
|||||
61 ctgctgtgctgaagaagaagctatgacgtctatctctgccaacattgagcag 120
|||||
85 CTCGTGTGCTGAAGGAAGCAAGCTATGAGCTCATCTGCTATCTGGCCAACTTGGCCAG 144
|||||
121 aagaagaagctcgaagaagcagaagaagcactgaagcttgaggcccaaaagtgttc 180
|||||
145 AAGGAAGCTTCGAGGAAGCCAGGAAGCAAGCACTGAAGCTTGGGGCCAAAGGTTC 204
|||||
181 attgagatgtcagcagaagattgttgagagatctatctgcccgcacaccagtcagc 240
|||||
205 ATTGAGATGTTCAGCAGGAGATTGTGTGAGAGATTCATCTGTGCGCCGCAATCCAGC 264
|||||
241 gcaatgtatgagagcagctacccctctggaacccctctctgcaagccctgcatcgcgc 300
|||||
265 GCACGTATGAGAGACCGCTACCTCTGCGGCACTCTCTTGGCAAGCCCTGATCGCCCGC 324
|||||
301 aacaagaaggaagaatccgcaagcagaagggcgaagatgtgtcccaagcgcgcacagga 360
|||||
325 AAACAAGTGAATGCGCCAGGAGGAGGCGCCCAAGTATGTGTCCACGCGCCACAGGA 384
|||||
361 aagggaagcagctcagctcagcttctgagctcagctcagctcagctcagctcagctcag 420
|||||
385 AAGGGGAACGATCAGAGCTCGGTTTGAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 444
|||||
421 gtcattgtctccctggaagatgctcgtgaattctacaacccggttcaagggccggaatgagctg 480
|||||
445 GTCATTTGCTCCCTGGAGGATGCTGAATTTTCAACCCGTTCAAGGGCGCAATGAGCTG 504
|||||
481 atgagtgatcgaagaagaacacaggaattcccaatccggtcactcccaagaacccgttgagc 540
|||||
505 ATGGAGTACGCAAGCAACAGGGGATTCCTCCATCCCGGTCACTCCCAAGAACCCGTGAGC 564
|||||
541 atgagtgagaacctcaatgacatcaagctcagagcttgaaatcctctgagagaaccccaagaac 600
|||||
565 ATGGATGGAACCTCATCTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 624
|||||
601 caagagctcagagctcctacaggaagacccaggaacccagcaagaagccccaacacccct 660
|||||
625 CAAGGCGCTCCAGGCTCTTACACGAAACCCAGGACCCAGCCAAAGCCCAAAACCCT 684
|||||
661 gacattctcgaagatcgaattcaaaa - aagggtcctctgtgaaagtgaacacgtlcaag 718
|||||
685 GACATTCTCGAGATCGAGTCAAAAACACAGGGGCTCCCTGTGAAGTGCACCAAGCTCAAGG 744
|||||
719 atggaac 778
|||||
745 ATGGACCAACCA - CAGACCTCTCTTGGAGCTTCTCATGTACTGAACCAAAATCTCCGGGCA 803
|||||
779 agcagcggttgagccgttatgtacatcgttgagagaacccgttcatttgaaatgagtcaccag 838
|||||
804 AGCATGGAGTGGGCGGTATTGACATCTGAGAACCCCGCTTATTGTGAATGACGTCGCCAG 863
|||||
839 gtaatacagagacc 853
|||||
864 GTATCTACGAGACC 878
|||||

RESULT 2

B1258834 878 bp mRNA EST 17-JUL-2001
LOCUS 602969646P1 NIH_MGC_112 Homo sapiens cdna clone IMAGE:5109251 5',
DEFINITION mRNA sequence.
ACCESSION B1258834
VERSION B1258834
KEYWORDS B1258834.1 GI:14815578
EST.

SOURCE

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 878)
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

COMMENT

Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1264 row: k column: 12
High quality sequence stop: 827.

FEATURES

source

1..878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5109251"
/clone_1ib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"

/note="Organ: cervix; Vector: PCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 225 a 250 c 246 g 157 t

ORIGIN

Query Match 65.4%; Score 809.8; DB 11; Length 878;
Best Local Similarity 98.6%; Pred. No. 4e-183;
Matches 869; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

8 gcaaaagctccgtgtgtctgacctacagtcgagcgtggaacccctgtcatcctgtgt 67
|||||
1 GCAAAAGGCTCGTGTGCTTGCTTACAGTGGCGGCTGGACACCTGTGCTATCTCTGTGT 60
|||||
68 gctgaaggaacaagcgtatgacgtcatctgtcatctgtgccaacattgagcagaagaag 127
|||||
61 GCCTGAAGGAACAAGGCTATGACGTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 120
|||||
128 acttcgaggaagcagaagaagaagcactgaagcttgaggccaaaaggtgtcatcattgag 187
|||||
121 ACTTCAAGGAAGCCAGGAAGGCACTGAAGCTTGGGCCCAAAAAGGTTTCATTATGAGG 180
|||||
188 atgtcagaagaggttctgaaagatctatctgagcagccatccagtlcagaacgacgt 247
|||||
181 ATGTCAAGAGAGAGTTTGTGAGAGATTTCATCTGCGGCGCAATCCAGTCCAGCGACCTGT 240
|||||
248 atgaagacgctacactctctgagcaactctctgcaagcgcctgcatcgccgcaaaagaag 307
|||||
241 ATGAGGACCCCTACTCTCTGCGCACCTCTTCCAGGCGCTGCAATCGCCCAAAACAAG 300
|||||
308 tggaaatcgccacagcgggaggggaggaagatgtgtgtcccaagcgcgcacagaagaagaag 367
|||||
301 TGGAAATCGCCACAGCGGAGGGGCCAAGATATGTGTCCACAGCGCCACAGGAAGAGGA 360
|||||
368 acgatacagctcgttctgagctcagctcagctcagctcagctcagctcagctcagctcagctcag 427
|||||

Db 361 AGATCAGTCCGGTTGAGCTACGCTTACTGAGCCGCCAGATAAGTCAATG 420

Qy 428 CTCCTGAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 487

Db 421 CTCCTGAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 480

Qy 488 ACCTGAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 547

Db 481 ACCTGAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 540

Qy 548 AATACCTGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 607

Db 541 AATACCTGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 600

Qy 608 CTCCTGAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 667

Db 601 CTCCTGAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 659

Qy 668 TCGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 726

Db 660 TCGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 719

Qy 727 ACCCTGAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 786

Db 720 ACCCTGAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 779

Qy 787 GTGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 846

Db 780 GTGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 837

Qy 847 GAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 887

Db 838 AAGAGATGCTGATCTACCAACCGTGTCAAGGGCCCAATGATGATGAGT 877

RESULT 3

LOCUS B1256559 838 bp mRNA EST 17-JUL-2001

DEFINITION 602976921F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5116133 5'

ACCESSION B1256559

VERSION B1256559.1 GI:14811085

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 838)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaab@remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L14M1282 row: j column: 06

High quality sequence stop: 810.

location/Qualifiers

1. 838

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NIH_MGC_12"

/lab_host="cervical carcinoma cell line"

/lab_host="DH10B"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; Notif: Site 2; Sali; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.4 kb. Library prepared by Life

BASE COUNT 212 a 237 c 239 g 150 t

ORIGIN

Query Match 64.8%; Score 803.2; DB 11; Length 838;

Best Local Similarity 99.4%; Pred. No. 1,5e-181;

Matches 827; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 3 gtcagcaaggctctcgtggtctgagctacagtggtgagctgagacacctgctatcct 62

Db 1 gtcagcaaggctctcgtggtctgagctacagtggtgagctgagacacctgctatcct 60

Qy 63 cgttgagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagct 122

Db 61 cgttgagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagct 120

Qy 123 ggaagctctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagct 182

Db 121 ggaagctctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagct 180

Qy 183 tgaagctctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagct 242

Db 181 tgaagctctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagct 240

Qy 243 actgtagagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 302

Db 241 actgtagagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 300

Qy 303 acaagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 362

Db 301 acaagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 360

Qy 363 ggaagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 422

Db 361 ggaagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 420

Qy 423 catgctcctcgtgagcaagcaagctgagctgagctgagctgagctgagctgagctgagct 482

Db 421 catgctcctcgtgagcaagcaagctgagctgagctgagctgagctgagctgagctgagct 480

Qy 483 ggaagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 542

Db 481 ggaagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 540

Qy 543 ggaagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 602

Db 541 ggaagctgagcaagcaagctgagctgagctgagctgagctgagctgagctgagctgagct 600

Qy 603 agcagctcagctcctcctacagcaagcaagcaagcaagcaagcaagcaagcaagcaagct 662

Db 601 agcagctcagctcctcctacagcaagcaagcaagcaagcaagcaagcaagcaagcaagct 660

Qy 663 catctcagctcagctcctcctacagcaagcaagcaagcaagcaagcaagcaagcaagct 722

Db 661 catctcagctcagctcctcctacagcaagcaagcaagcaagcaagcaagcaagcaagct 720

Qy 723 caccagctcagctcctcctacagcaagcaagcaagcaagcaagcaagcaagcaagcaagct 782

Db 721 caccagctcagctcctcctacagcaagcaagcaagcaagcaagcaagcaagcaagcaagct 779

Qy 783 tggcgtgagctcagctcctcctacagcaagcaagcaagcaagcaagcaagcaagcaagct 834

Db 780 tggcgtgagctcagctcctcctacagcaagcaagcaagcaagcaagcaagcaagcaagct 830

RESULT 4

LOCUS B1257913 875 bp mRNA EST 17-JUL-2001

DEFINITION 602970844F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5110307 5'

ACCESSION B1257913

VERSION B1257913.1 GI:14813752

```

Best Local Similarity
63.0%; Score 780.6; DB 11; Length 815;
99.0%; Pred. No. 3.7e-176;

```



```

|||||
Db 541 GGAGAACGGCTTCATTGATGATAGTCCCGAGGTATCTACGAGACCCCGACAGGACACCA 600
Qy 866 tcccttaacacatgctcatttaagacatcgagagccttcacatgagccgggaagtgcacaaa 925
Db 601 TCCCTTACCATGCTCATTTAGACATCGAGGCGCTTCACCATGAGCCGGGAGAGTGGCCAAAA 660
Qy 926 tcaaacagagcctggcttgaattgctgaagctgctgataccgagcttaagccttaagc 985
Db 661 TCAACACAGGCTGGGCTTGAATTTGCTGAGCTGTGTATACCGGTTTCTGGCCAAAGC 720
Qy 986 ctgaatgtgaattgtccgcacatgacatc--gccaaatcccaagagagcagatgag--aaagga 1042
Db 721 CTGAGTGTGAATTTGTCGGCCGACATCGCCGCAAGGTCCGAGGAGGAGGAGGACGCGGA 780
Qy 1043 aagtcgaaggtgcgtctcctcaagagcc--aagtgtaacatcctgcgcggagatgctccac 1099
Db 781 AAGTGCAGGTGTCCCTCCCTCAAGGGGCGCAGGTGTTCATTCCTCGCGCGGAGATCCGCC 840
Qy 1100 tgcctctctacaatgagagagagagagagagagagagagagagagagagagagagagag 1153
Db 841 CTGGTCTCTCTTACATGAGGAGAGGCTGTGTGACCTTGTGACGTGACGCGGTGATCTATGAG 900
Qy 1154 caactgatagcacacggtatcatcaacatcaatcctcctcaagcctgaaaggaat 1204
Db 901 CCACGTGATTCGCCCGGGTCATCAACACATCATTTCCCGCAGGTGAGAGAGAT 951

```

```

RESULT 7
Bg828136 905 bp mRNA EST 22-MAY-2001
LOCUS 602753522F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4906132 5',
DEFINITION mRNA sequence.
ACCESSION Bg828136
VERSION Bg828136.1 GI:14175723
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 905)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCX1808 row: 1 column: 05
High quality sequence stop: 790.

```

FEATURES

source

```

1..905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4906132"
/clone_11b="NIH_MGC_17"
/tissue_type="thadomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pOTB7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(6). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT 227 a 265 c 253 g 160 t
ORIGIN
```

```

Query Match 62.3% Score 772.4 DB 11: Length 905;
Best Local Similarity 97.1% Pred. No. 3.5e-174;
Matches 851; Conservative 0; Mismatches 16; Indels 9; Gaps 6;

Qy 1 atcgcagcaaaagctcgcgtgcttcgcttaacagtcgagcgccttgacacctcgtgac 60
Db 27 ATGCCAGCAAAAGGCTCCGTTCTGGCTTACAGTGGCGGCTGGACACCTGTGCATC 86
Qy 61 ctgctgtgctgaagaaagcaagcgtatgacgtcattgctctatctggtgcaaatgtgccag 120
Db 87 CTCCTGTGCTGGAAGGAACAAGGCTATGACGTCTATTCCTGTGCGCAACATTTGGCCAG 146
Qy 121 aagaaagactctgagaaagcagaagaaagcaactgaagcttgaggccaaa--aagtgct 179
Db 147 AAGGAAGACTTTCAGAGAACCCAGAGAAAGGCACTGAAGCTTGGGCGCAAAAGGTGT 206
Qy 180 catgaagatgctcagcaagagagcttgatgagagatcatctggtgcgagcatcagtcag 239
Db 207 CATGAGAGATGTGACGAGGAGTTTGTGAGAGTTTCATCTGCGCCGATCCAGTCCAG 266
Qy 240 cgcactgtatgagagacgcgtactcctctggtgacactctctgccaagccttgacatgc 299
Db 267 CGCAGTGTATGAGAGACCGCTACTCTCGGCGACCTCTCTTCCAGGCGCTGCAATCGCCG 326
Qy 300 caaacgaatggaatgcgcccagcagagagagagagagagatgctgtcccaagcgcccaag 359
Db 327 CAACAGAGTGAATTCGCCCAAGCGGAGGGGCGCAAGTATGTCTCCACGCGCCACAG 386
Qy 360 aagaagaaagatcagatcagctgcttgatgagctcagctgactacatcagtcgccccagataa 419
Db 387 AAAGGGAAGAGATCAGGTCCGGTTTGTAGCTCAGTGTCTACTGCTGAGCCGCCAGATAA 446
Qy 420 ggtcattgctccctcgtgagagatgctgataatctcaaacagcgttcaagagcgccgaat 479
Db 447 GGTATTGCTCCCTCGAGAGATGCTGAATTCACAAACGGTTCAGAGGCGCGCATATACCT 506
Qy 480 gatgagatgagcaaaagcaacagagatcccatcccggtacatcccaagaacccgtgag 539
Db 507 GATGAGATGCGCAAGACACGAGGATTCCTCATCCGCTACCTCCCAAGAACCCGTGAG 566
Qy 540 catgatatgaacatcactgacatcaagctcaagagctgaaatccctgagaaaccccaagaa 599
Db 567 CATGATGAGAACTCTCATGACATCAGCTACGAGGCTGGAATCTGAGAAACCCCAAGAA 626
Qy 600 ccaagcgctcctcagctgtctcaccagaaagccagagcccaagagcccccagacccc 659
Db 627 CCAAGCGCTCCAGGCTCTTACACGAGAGACCCAGGACCCAGCCAAAGCCCCCAA -ACCC 685
Qy 660 tgaatctcagatcagatcagatcacaagagaggtccctgtgaaggtgaagcaagtcacaag 719
Db 686 TGACATTTCTGAGATGAGTTCGATCAAAACAAAGGGGTCTGTAGAGTGACCAAGCTCAAG 745
Qy 720 tggcacaacccaacagacatcctctgagctctctcaatgataagcaagatgcgagc 779
Db 746 TGGCACACACACAGCA---CTCCTTGAGGCTCTTATCTATCTACCTAAGCAATCCCGGCA 802
Qy 780 gcatgagcgtgagcgtatgataatcgtgagagacccgttcaatgagaaagtcgagcag 839
Db 803 GCATGGGCTGGGCGGTATTGACATCGTGGAGAACCG--TTATATGGAAATGAAGTCC-AGG 860
Qy 840 tatctagagaccccaagcagacacatccttaca 875
Db 861 GATCTAGGAAMCCCA--TTGGAGCATTCCTTACCA 894

RESULT 8
Bi259920 918 bp mRNA EST 17-JUL-2001
LOCUS 60296896F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5108463 5',
DEFINITION mRNA sequence.
ACCESSION Bi259920
```

VERSION B1259920.1 GI:14817719
 KEYWORDS EST.
 SOURCE Human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 918)
 AUTHORS NIH-MGC
 TITLE NIH-MGC http://mgi.ncl.nih.gov/
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM1262 row: j column: 16
 High quality sequence stop: 834.

FEATURES
 source
 1. 918
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5108463"
 /clone_lib="NIH_MGC_12"
 /tissue_type="Cervical carcinoma cell line"
 /note="Organ: cervix; Vector: PCMV-SPORT6; Site: J; Note:
 Site_2: Salt: Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.4 Kb. Library prepared by Life
 Technologies"

BASE COUNT 241 a 266 c 242 g 169 t
 ORIGIN

Query Match 62.1%; Score 768.8; DB 11; Length 918;
 Best Local Similarity 95.6%; Pred. No. 2.5e-173;
 Matches 876; Conservative 0; Mismatches 32; Indels 8; Gaps 8;

310 gaatgcccag-cggagggggccagatgtgtcccaag-9cgccacag-gaaagg 366
 1 GAAATGCCCGGAGGGGGGCGCAAGTATGTCTCCACGTTGTCACAGAGGAAGGGG 60
 367 aacgacagtcgggttgagctgactgactgccccccagataaagtcatt 426
 61 AACGATCGGTCGGTTGAGCTCAGCTGCTACTGCGCCCGCCAGATAAAGTCATT 120
 427 g-cctccctggaagagtccttaattcaacacgggtcaaggcgcaatgacctgata 485
 121 GCTCCCTGGAGAGTCTGGAATTTCTACAAACCGGTTCAAGGGCCGCAATGACCTGATGGA 180
 486 gtaagcaaacacacagggatcccatcccggtcactcccaagaacccggagatgga 545
 181 GTACGCAAGCAACAGGAGATCCATCCCGTCACTCCCAAGACCCGGGAGCATGGA 240
 546 tggagacccatgacatcagctagagtgagatcttgagaaaccccaagaacacac 605
 241 TGGAAACTCAATGACATCAGCTGAGGCTGGAATCTTGGAAGACCCCAAGAACCAAG 300
 606 gctccaggtctctacaagaagaccagagcccaagcccaaaccccaaccccgat 665
 301 GCCTCAGGTCCTACACGAAGACCGAGCCAGCAAGCCCAACCCCTGACAT 360
 666 tctgagatgagttcaaaaaagggtccctgtgaagtgagcaacagtcgaagatggac 725
 361 TCTGAGATGAGTTCAAAAAAGGGTCCCTGTGAAGGTGACCAACGTCACAGATGGAC 420
 726 caccaccagacacctctgagctctatgacctgaacgaagtcggcggaagcatg 785
 421 CAGCACCAACACCTCTTGAGACTTCTCATGACTGAGCAAGTCCGGGCAAGCATGG 480

QY 786 cgtggccgtattgacatcgtgaggaacggtccattggaatgaagtcocaggatarta 845
 Db 481 CGTGGCCGTTGATCGTGGAGAACCGCTTCATTTGAATGAAGTCCGAGGTATCTA 540
 QY 846 cgaagcccccagagagacccatcccttacatgcatatgacatgagagccttaacat 905
 Db 541 CGAGACCCCAAGAGGACACCTTCCTTACCATGCTCATTTAACAATCGAGGCTTCAACAT 600
 QY 906 ggaacccggaagtgagcaaaatacaacaagcctggcttga-aattgctgaagtgat 964
 Db 601 GGACCGGGAAGTGGCAAAATCAACAGGCTTGGCTTGAACAATTTGCTGAGACTGTGT 660
 QY 965 ataccggttaagcgcctgagctgagtgatgaattgtccgcacatgcatcgcg-aagtc 1023
 Db 661 ATACCGGTTTTCGACACACGCTGAGTGTGAATTTGTCGCCCATCTCATCGCAAAAGTCC 720
 QY 1024 cagagagagtggaaggaagtgagc-agtgctgcctcccaaggagcagtgatcatcct 1082
 Db 721 CAGGACGAGTGGAAAGGAAAGTCAAGGTCCTGCTCAAGGGCCAGGTGTAACATCC 780
 QY 1083 cggccggagagtcgccactgtctctcaaatgagagcgtgtgagcatgagcgtgag 1142
 Db 781 TCGCCGGGAATCCCATGTCCTCTCAATGAGAGCTGGAGCATGAACTGCCGA 840
 QY 1143 tgattatg-agcaactgagccacccgggttcatcaatcaatctccctgaagtgag 1201
 Db 841 TGATCATGAAGCCAAATGAATGCCAACCGAGTCATCATCAATTCCTCCTCAGTGAAG 900
 QY 1202 aatatactgctcca 1217
 Db 901 ACTATCATCGACTCCA 916

RESULT 9

BE378581
 LOCUS 952 bp mRNA
 DEFINITION 60123639P1 NIH_MGC_44 Homo sapiens CDNA clone IMAGE:3609032 5',
 mRNA sequence.
 ACCESSION BE378581
 VERSION BE378581.1 GI:9323946
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 952)
 AUTHORS NIH-MGC
 TITLE NIH-MGC http://mgi.ncl.nih.gov/
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM262 row: n column: 09
 High quality sequence stop: 766.
 Location/Qualifiers
 1. 952
 Location/Qualifiers

FEATURES
 source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3609032"
 /clone_lib="NIH_MGC_44"
 /tissue_type="Endometrium, adenocarcinoma cell line"
 /note="Organ: uterus; Vector: pOTB7; Site: 1; XhoI: Site_2:
 EcoRI; CDNA made by oligo-dT priming. directionally
 cloned into Rocti/XhoI sites using the following 5'
 adaptor: GGCAAGAG(6). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of

	Score	DB	Length
Query Match	61.2%	758.4	952
Best Local Similarity	96.2%	7.7e-171	
Matches	820	Mismatches	26
		Indels	6
		Gaps	4

RESULT
BI085601

```

/organism="taxon:9606"
/db_xref="IMAGE:5013131"
/clone_1fb="NIH.MGC_98"
/tissue-type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI library constructed by Ling Hong
 adaptor: GGCACGAG(G). Gerald M. Rubin (University of
 California, Berkeley) using zap-CDNA synthesis kit
 (stratagene) and Superscript II RT (life Technologies)."
Note: this is a NIH.MGC library."

```

BASE COUNT	223 a	252 c	253 g	153 t
ORIGIN				
Query Match		60.8%;	Score 753.2;	DB 11; Length 881;
Best Local Similarity	96.2%;	Pred. No. 1.3e-169;		
Matches 825; Conservative		0; Mismatches 28;	Indels 5;	Gaps 5;
Db	1 atgtcagcaaaaggtctcggtgtgtcttctgtgctacagtgtgcgtgcgccttggaaacctgtgtcatc 60			
	23 ATGTCCGCAAAAGGCTCCGTGTTCTGTGGCCACAGTGGCGGCTTGACACACTCTGTGATC 82			
Oy	61 ctgtgtgtgtgtgaagaaacaggtctatagcttcttcattcctatctgtgccaattgtgccag 120			
	83 CTCGTGTGTGCGTGAAGAGACAAAGCATACGCTATACGCTATTGCTCATCTGTGGCCAACTTTGGCCAG 142			
Db	121 aagaaagactctcgaggaagaagcagaagaagcacttgaagctttgggccaagaagtgtctc 180			
Oy	143 AAGGAAGACTTTCGAGGAGCCAGAACGAAGACACTGAAGCTTGGGGCCAAANAAGTGTTC 202			
Db	181 attgagagatgcacagaggaagtttgtgtgagaggtctcatcttgcgcgcatctcagtcacagc 240			
Oy	203 ATTGAGGATGTCACGACGAGGATTTGTGGAGGATTCATCTGCGGCGCATTCAGTCTACAC 262			
Db	241 gcaactgtatgagacgcgttaactctctctgtggagacctctctgtccaagccctgcatctgcgcgcg 300			
Oy	263 GCACGTGTATGAGACCGCTACTCTCTGTGGACCTCTCTTGGCCAGGCCCTTGATGCCCCC 322			
Db	301 aaacaagtgtgaatctgccacgcgagaggggtccaagtatgtgtccacagtcgacacagaga 360			
Oy	323 AAACAAGTGAATTCGCCACAGGGAGGGGCCAAGATGATGTCTCCACAGGGCCACAGGA 382			

Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLND at: <http://image.llnl.gov>
 Plate: LAM11280 row: 1 column: 06
 High quality sequence spot: 815.
 Location/Qualifiers

```

LOCATION/xxxxx=
1. . 898
/organism="Homo sapiens"

```

229 a	251 c	255 g	257
-------	-------	-------	-----

Query Match	Score	DB	Length
Best Local Similarity	748.6;	11;	898;
Matches	860;	Conservative	0;
		Mismatches	19;
		Indels	10;
		Gaps	9;

[illegible]

QY	361	aaggggaacgaataagatgcgcggtttgagtcacgaatgctactcaactgcccccaataaa-a	419
Db	377	AAGGGGAACGATCAGTCCGGTTTGAGCTCACCTCTACTCTGGCCCCCAGATMAA	436
QY	420	ggtcattgtccctcggagagatgctcctaattcacaacccggtttcaaggccgaatgact	479
Db	437	GGTCATTGCTCCTCGAGGATGGCTGAATTTCACAACCGGTTCAAGGGCCCGCAATGACTT	496
QY	480	gatagagtagcagaagaacaacggaattccattcccggttactccagaaccggttgag	539
Db	497	GATGAGTATAGCAAAACAACGGGATTCCTATCCCGGTCACTCCCAAGAACCCGTGGAG	556
QY	540	catgatagaaacctatgcacatcaagctcagaagcttgaatctctgtgaagaccacaaga	599
Db	557	CATGATGATGAACCTCATGTCATGAGTACGAGGCTGGAATCTCGAGAA-CCCAAGAA	615
QY	600	ccaagcgctccagctctctacagaagaaccagagaccagacaaagccccaacacccc	659
Db	616	CCAAACGGCTCTCAGGTCTTACACGMAA-CCAGAGACCCAGCCAAAGCCCCAAGA-CCC	673
QY	660	tgaactcttcgagatcgagttcaaaaagggtgtccctgttgaagctgaaccaatgctcaaga	719
Db	674	TGACATTCCTCGAGATCGAGTCTCAAAAAGGGGTCCCTGTGAAGTTGAACAAGTCAAGA	733
QY	720	tggacacacacacacagacaccttgaagctcttcaatgtaacctgaagaagtc-aggagca	778
Db	734	TGGCAC--ACCAACAGAAACTCTTGAAGGCTTTCATGTACTGTAGCAAGTCAAGGGGCA	791
QY	779	agaatggcgttgggcgcgtatgtaatatgttgaagaaccgttcatatgtaatgaagtcagg	838
Db	792	AGCATGGCTGGGGCGTATGACATCTCTGGAGAAACGGTATTATGGACACTGAAT-CCAG	850
QY	839	gtactacagacccacagagaccatctcttacaatgtaacttattaga	887
Db	851	GTATCTACGAGA-CCAGAGAGGGCAACTTCTTACCATCTCATTTTGA	898

RESULT	13	17-JUL-2001
BI258628	851 bp	EST
LOCUS	602969566F1	NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5109064 5'
DEFINITION	mRNA sequence.	
ACCESSION	BI258628	
VERSION	BI258628.1	GI:14815170

SOURCE
CONTACTS

SOURCE	ORGANISM
Human.	Homo sapiens
Notazoa.	Chordata; Vertebrata; Euteleostomi

APPENDIX

REFERENCE

AUTHORS

Title

JOURNAL

COMMENT

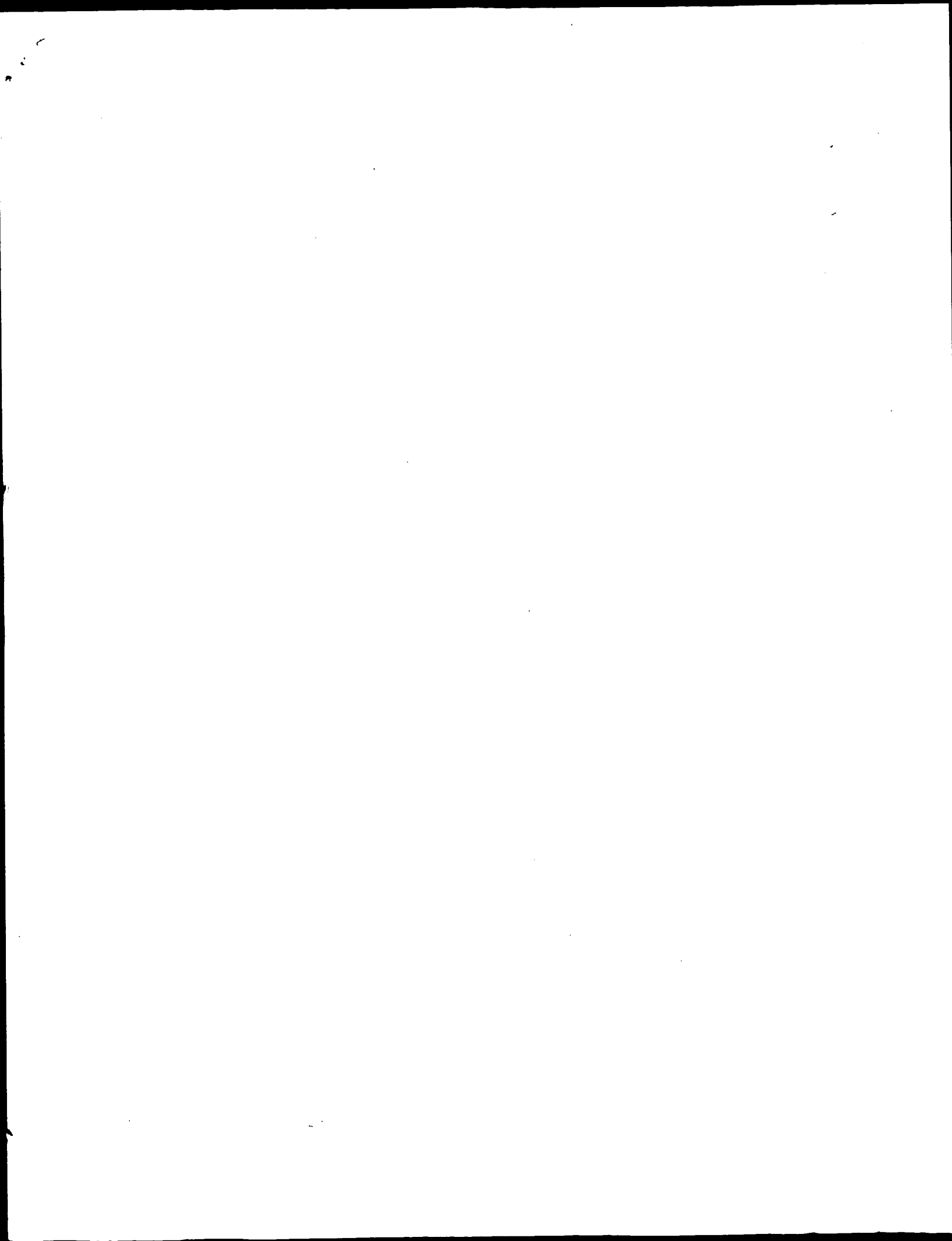
FEATURES

Source

[REDACTED]

Wed Feb 13 07:36:09 2002

us-09-775-693-1.rst



Davis, M.
09/775693
Seq. ID 1 w/ Interf
Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:46:27 ; Search time 2570.39 Seconds
(without alignments)
8448.217 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1 atgtccagcaagagctcgt.....gcaaggtcactgccaatag 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 17159718 seqs, 876320856 residues

Total number of hits satisfying chosen parameters: 34319436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending-Patents_NA_Main:*

1:	/cgn2_6/ptodata/2/pna/US0611.COMB.seq.*
2:	/cgn2_6/ptodata/2/pna/US0612.COMB.seq.*
3:	/cgn2_6/ptodata/2/pna/US0613.COMB.seq.*
4:	/cgn2_6/ptodata/2/pna/US0614.COMB.seq.*
5:	/cgn2_6/ptodata/2/pna/US0615.COMB.seq.*
6:	/cgn2_6/ptodata/2/pna/US0616.COMB.seq.*
7:	/cgn2_6/ptodata/2/pna/US0617.COMB.seq.*
8:	/cgn2_6/ptodata/2/pna/US0618.COMB.seq.*
9:	/cgn2_6/ptodata/2/pna/US0619.COMB.seq.*
10:	/cgn2_6/ptodata/2/pna/US0620.COMB.seq.*
11:	/cgn2_6/ptodata/2/pna/US0621.COMB.seq.*
12:	/cgn2_6/ptodata/2/pna/US0622.COMB.seq.*
13:	/cgn2_6/ptodata/2/pna/US0623.COMB.seq.*
14:	/cgn2_6/ptodata/2/pna/US0624.COMB.seq.*
15:	/cgn2_6/ptodata/2/pna/US0625.COMB.seq.*
16:	/cgn2_6/ptodata/2/pna/US0626.COMB.seq.*
17:	/cgn2_6/ptodata/2/pna/US0627.COMB.seq.*
18:	/cgn2_6/ptodata/2/pna/US0628.COMB.seq.*
19:	/cgn2_6/ptodata/2/pna/US0629.COMB.seq.*
20:	/cgn2_6/ptodata/2/pna/US0630.COMB.seq.*
21:	/cgn2_6/ptodata/2/pna/US0631.COMB.seq.*
22:	/cgn2_6/ptodata/2/pna/US0632.COMB.seq.*
23:	/cgn2_6/ptodata/2/pna/US0633.COMB.seq.*
24:	/cgn2_6/ptodata/2/pna/US0634.COMB.seq.*
25:	/cgn2_6/ptodata/2/pna/US0635.COMB.seq.*
26:	/cgn2_6/ptodata/2/pna/US0636.COMB.seq.*
27:	/cgn2_6/ptodata/2/pna/US0637.COMB.seq.*
28:	/cgn2_6/ptodata/2/pna/US0638.COMB.seq.*
29:	/cgn2_6/ptodata/2/pna/US0639.COMB.seq.*
30:	/cgn2_6/ptodata/2/pna/US0640.COMB.seq.*
31:	/cgn2_6/ptodata/2/pna/US0641.COMB.seq.*
32:	/cgn2_6/ptodata/2/pna/US0642.COMB.seq.*
33:	/cgn2_6/ptodata/2/pna/US0643.COMB.seq.*
34:	/cgn2_6/ptodata/2/pna/US0644.COMB.seq.*
35:	/cgn2_6/ptodata/2/pna/US0645.COMB.seq.*
36:	/cgn2_6/ptodata/2/pna/US0646.COMB.seq.*
37:	/cgn2_6/ptodata/2/pna/US0647.COMB.seq.*
38:	/cgn2_6/ptodata/2/pna/US0648.COMB.seq.*
39:	/cgn2_6/ptodata/2/pna/US0649.COMB.seq.*
40:	/cgn2_6/ptodata/2/pna/US0650.COMB.seq.*
41:	/cgn2_6/ptodata/2/pna/US0651.COMB.seq.*
42:	/cgn2_6/ptodata/2/pna/US0652.COMB.seq.*
43:	/cgn2_6/ptodata/2/pna/US0653.COMB.seq.*

Result

No.	Score	Query Match	Length	DB	ID	Description
1	1239	100.0	1239	30	US-09-775-693-1	Sequence 1, Appl
2	1232.6	99.5	1622	22	US-09-577-410-7427	Sequence 7427, Ap
3	1232.6	99.5	1656	17	US-09-338-425-2021	Sequence 2021, Ap
4	1232.6	99.5	1656	25	US-09-652-126-9702	Sequence 9702, Ap
5	1232.6	99.5	1656	25	US-09-652-816-9327	Sequence 9327, Ap
6	1232.6	99.5	1656	27	US-09-698-010-14283	Sequence 14283, A
7	1232.6	99.5	1656	27	US-09-698-012-8607	Sequence 8607, Ap
8	1232.6	99.5	1656	28	US-09-710-281-5043	Sequence 5043, Ap
9	1232.6	99.5	1656	28	US-09-716-920-1274	Sequence 1274, Ap
10	1232.6	99.5	1656	29	US-09-726-175-2581	Sequence 2581, Ap
11	1232.6	99.5	1656	29	US-09-726-787-2755	Sequence 2755, Ap
12	1232.6	99.5	1656	29	US-09-726-790-2021	Sequence 2021, Ap
13	1232.6	99.5	1755	30	US-09-732-630-4186	Sequence 4186, Ap
14	1232.6	99.5	1755	30	US-09-770-173-2339	Sequence 2339, Ap
15	1232.6	99.5	2057	30	US-09-760-475-983	Sequence 983, Ap
16	1232.6	99.5	2810	1	PCT-US00-265248-1503	Sequence 1903, Ap
17	1232.6	99.5	2812	32	PCT-US00-059888-654	Sequence 654, Ap
18	1232.6	99.5	2812	32	US-09-925-300-654	Sequence 654, Ap
19	1231	99.4	1571	56	US-60-239-841-86	Sequence 86, Appl
20	1221.6	98.6	1836	60	US-60-278-258-389	Sequence 389, Appl
21	1220	98.5	1808	17	US-60-164-285-7559	Sequence 7559, Ap
22	1200	96.9	1808	17	US-09-359-922-639	Sequence 639, Appl
23	1200	96.9	1808	17	US-09-359-922-639	Sequence 639, Appl
24	1199.6	96.8	1828	16	US-09-340-620-13594	Sequence 13594, A
25	1199.6	96.8	1828	16	US-09-340-620-13594	Sequence 13594, A
26	1199.6	96.8	1828	31	US-09-898-888-13594	Sequence 13594, A
27	1199.6	96.8	1828	31	US-09-898-888-13594	Sequence 13594, A
28	1093	88.2	164833	54	US-60-216-770-49	Sequence 49, Appl
29	1093	88.2	164833	54	US-60-216-770-49	Sequence 49, Appl
30	1061	85.6	4886	57	US-60-243-468-420	Sequence 420, Appl
31	1058	85.4	191307	54	US-60-216-770-45	Sequence 45, Appl
32	1058	85.4	191307	54	US-60-216-770-45	Sequence 45, Appl
33	1031.8	83.3	3589	53	US-60-212-664-168	Sequence 168, Appl
34	1030	83.1	5018	57	US-60-207-216-1442	Sequence 342, Appl
35	1012.4	81.7	104925	54	US-60-243-468-528	Sequence 528, Appl
36	913.2	72.0	1541	60	US-60-212-664-284	Sequence 284, Appl
37	892	73.7	2252	54	US-60-278-258-387	Sequence 387, Appl
38	828	66.8	152425	54	US-60-212-664-165	Sequence 630, Appl
39	742.8	60.0	1871	1	US-60-212-664-165	Sequence 165, Appl
40	731.6	59.0	3017	51	PCT-US01-08631-4443	Sequence 4443, Ap
41	731.6	59.0	3017	51	US-60-189-499-92	Sequence 92, Appl
					US-60-189-499-92	Sequence 142, Appl

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

C	42	730	58.9	3017	51	US-60-185-362-59	Sequence 59, Appl
C	43	730	58.9	3017	51	US-60-185-362-126	Sequence 126, App
44	723.8	58.4	953	57	US-60-243-468-1942	Sequence 1942, Ap	
45	718.4	58.0	42519	54	US-60-216-770-149	Sequence 149, App	

ALIGNMENTS

RESULT 1

US-09-775-693-1

Sequence 1, Application US/09775693

GENERAL INFORMATION:

APPLICANT: Clark, Mike

APPLICANT: Holtsberg, Frederick Wayne

APPLICANT: Ensor, Charles Mark

TITLE OF INVENTION: Methods For Predicting Susceptibility Of Patients To Arginine Def

FILE REFERENCE: Therapy

CURRENT APPLICATION NUMBER: US/09/775,693

CURRENT FILING DATE: 2001-02-02

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1

LENGTH: 1239

TYPE: DNA

ORGANISM: Homo sapiens

US-09-775-693-1

Query Match 100.0%; Score 1239; DB 30; Length 1239;
 Best Local Similarity 100.0%; Pred. No. 1,5e-295;
 Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	atgttcagcaaaagctccgtgtctctgctaccagtgccgctggagacactgtgcatc	60
DB	1	atgttcagcaaaagctccgtgtctctgctaccagtgccgctggagacactgtgcatc	60
QY	61	ctcgtgtgtgtggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag	120
DB	61	ctcgtgtgtgtggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag	120
QY	121	aaggaagacttcgaggaagcaggaaggaaggaaggaaggaaggaaggaaggaaggaag	180
DB	121	aaggaagacttcgaggaagcaggaaggaaggaaggaaggaaggaaggaaggaaggaag	180
QY	181	atgaggaatgtcagcaaggaaggttctgaggaaggaaggaaggaaggaaggaaggaag	240
DB	181	atgaggaatgtcagcaaggaaggttctgaggaaggaaggaaggaaggaaggaaggaag	240
QY	241	gcactgtatgaggaagcagcctcctcctcctcctcctcctcctcctcctcctcctcctc	300
DB	241	gcactgtatgaggaagcagcctcctcctcctcctcctcctcctcctcctcctcctcctc	300
QY	301	aaacaagtggaatcgcgcagcggaaggggccaagatattgtgccacgagcgccacagga	360
DB	301	aaacaagtggaatcgcgcagcggaaggggccaagatattgtgccacgagcgccacagga	360
QY	361	aaaggaagcagatcaggttcggttctgaggtcaggttcaggttcaggttcaggttcaggt	420
DB	361	aaaggaagcagatcaggttcggttctgaggtcaggttcaggttcaggttcaggttcaggt	420
QY	421	gtcattgtcctccctgaggaatcctcctcctcctcctcctcctcctcctcctcctcctc	480
DB	421	gtcattgtcctccctgaggaatcctcctcctcctcctcctcctcctcctcctcctcctc	480
QY	481	atggaatgaagcaagcaacaggaattccatcccggttactcccaagaagaccgttgagc	540
DB	481	atggaatgaagcaagcaacaggaattccatcccggttactcccaagaagaccgttgagc	540
QY	541	atggaatgaagcaagcaacaggaattccatcccggttactcccaagaagaccgttgagc	600
DB	541	atggaatgaagcaagcaacaggaattccatcccggttactcccaagaagaccgttgagc	600

QY	601	caagcgtccctcaggtctctctacacgaagacccaggaagcccaagcccaaacacccct	660
DB	601	caagcgtccctcaggtctctctacacgaagacccaggaagcccaagcccaaacacccct	660
QY	661	gacattctcgaatcaggttcaaaaaaagggttcccttgagagtgacaaagctcaagat	720
DB	661	gacattctcgaatcaggttcaaaaaaagggttcccttgagagtgacaaagctcaagat	720
QY	721	ggcaccacccacagactcctcttgaggtctctccttcaatgaactgaagcaagtcgggcaag	780
DB	721	ggcaccacccacagactcctcttgaggtctctccttcaatgaactgaagcaagtcgggcaag	780
QY	781	catggtctggtcgtatttgacatcgttgaggaacgccttcaatgaatgaagtcggaggt	840
DB	781	catggtctggtcgtatttgacatcgttgaggaacgccttcaatgaatgaagtcggaggt	840
QY	841	atctagagagcccccaggaagcaccatccttaccatgctcaatttagacatcgagccttc	900
DB	841	atctagagagcccccaggaagcaccatccttaccatgctcaatttagacatcgagccttc	900
QY	901	accatggaccgggaagtgccgcaaatcaaaccaagccttggttggaatttgcgtgagc	960
DB	901	accatggaccgggaagtgccgcaaatcaaaccaagccttggttggaatttgcgtgagc	960
QY	961	gtgtataccggtttagcgtcctagccttgaattgttcgcgcacatcgacatcgacag	1020
DB	961	gtgtataccggtttagcgtcctagccttgaattgttcgcgcacatcgacatcgacag	1020
QY	1021	tcccaggaagcagtggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag	1080
DB	1021	tcccaggaagcagtggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag	1080
QY	1081	ctcggccgggaagtcaccactgtctctcctacaaatggaagcgttgagcaatgagcgtg	1140
DB	1081	ctcggccgggaagtcaccactgtctctcctacaaatggaagcgttgagcaatgagcgtg	1140
QY	1141	gtgtataggaagcaactatgacacgggttcacatcaaatcaatcaatcaatcaatcaat	1200
DB	1141	gtgtataggaagcaactatgacacgggttcacatcaaatcaatcaatcaatcaatcaat	1200
QY	1201	gaatatacgtctcctcaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag	1239
DB	1201	gaatatacgtctcctcaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag	1239

RESULT 2

US-09-577-410-7427

Sequence 7427, Application US/09577410

GENERAL INFORMATION:

APPLICANT: Gutierrez-Ramos, Jose-Carlos

APPLICANT: Hodge, Martin

APPLICANT: Kingsbury, Gillian

APPLICANT: Mackay, Charles

TITLE OF INVENTION: Nucleic Acid Molecules Derived from

FILE REFERENCE: 5800-32

CURRENT APPLICATION NUMBER: US/09/577,410

CURRENT FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: US 60/135,632

PRIOR FILING DATE: 1999-05-24

PRIOR APPLICATION NUMBER: US 60/135,633

PRIOR FILING DATE: 1999-05-24

SOFTWARE: FastSeq for Windows Version 3.0

NUMBER OF SEQ ID NOS: 8991

SEQ ID NO 7427

LENGTH: 1622

TYPE: DNA

ORGANISM: Homo sapiens

US-09-577-410-7427


```

Db 545 gtcatctctccctgagagatgccttgatctacaacccgttcaaggccgcaatgacctg 604
QY 481 atgagatagcaagaagaacagggatctccatcccggtcactcccaagaacccgtggagc 540
Db 605 atggagtagcaagaagaacagggatctccatcccggtcactcccaagaacccgtggagc 664
QY 541 atgagatagcaagaagaacagggatctccatcccggtcactcccaagaacccgtggagc 600
Db 665 atgagatagcaagaagaacagggatctccatcccggtcactcccaagaacccgtggagc 724
QY 601 caagcgcctccagagctctacacgaagcccaagcccaagcccaagcccaagcccaagcc 660
Db 725 caagcgcctccagagctctacacgaagcccaagcccaagcccaagcccaagcccaagcc 784
QY 661 gacattctcgagatcgagttcaaaaagggtccctgtgaggttgaggtgaggtgaggtg 720
Db 785 gacattctcgagatcgagttcaaaaagggtccctgtgaggttgaggtgaggtgaggtg 844
QY 721 ggcacacacacacacacacacacacacacacacacacacacacacacacacacacac 840
Db 845 ggcacacacacacacacacacacacacacacacacacacacacacacacacacacac 904
QY 781 catggtctgggctgatttgacatctgagagacgccttcaattgaggtgaggtgaggtg 840
Db 905 catggtctgggctgatttgacatctgagagacgccttcaattgaggtgaggtgaggtg 964
QY 841 atctacagaagcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 900
Db 965 atctacagaagcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1024
QY 901 accatggagccgggaagtgagcaaaatcaaaagaagccctggggttgaatttgctgagctg 960
Db 1025 accatggagccgggaagtgagcaaaatcaaaagaagccctggggttgaatttgctgagctg 1084
QY 961 gtgtacacgggttaagcgttagccctgagtgaggttgcctgcacatgacgtgacgaag 1020
Db 1085 gtgtacacgggttctgagacagcccttgagtgaggttgcctgcacatgacgtgacgaag 1144
QY 1021 tcccaagagcagtgaggaaggaagtgaggtgaggtgaggtgaggtgaggtgaggtgag 1080
Db 1145 tcccaagagcagtgaggaaggaagtgaggtgaggtgaggtgaggtgaggtgaggtgag 1204
QY 1081 ctcggccgggagctcccaactgtctctcaaaatgaggaagtgaggtgaggtgaggtgag 1140
Db 1205 ctcggccgggagctcccaactgtctctcaaaatgaggaagtgaggtgaggtgaggtgag 1264
QY 1141 ggtgatttgagccaactgacgtgacccgggtgtcaatcaatcaatccccaagctgag 1200
Db 1265 ggtgatttgagccaactgacgtgacccgggtgtcaatcaatcaatccccaagctgag 1324
QY 1201 gaatacatgctccacagagcagagtgatctgcaaatag 1239
Db 1325 gaatacatgctccacagagcagagtgatctgcaaatag 1363

```

US-09-652-126-9702

Query Match 99.5%; Score 1232.6; DB 25; Length 1656;
 Best Local Similarity 99.7%; Pred. No. 6.1e-294;
 Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 atgtcaacacaaagctccgtgtctctgctacagatggcggtctgacacactgtgacac 60
Db 125 atgtcaacacaaagctccgtgtctctgctacagatggcggtctgacacactgtgacac 184
QY 61 ctctgtgtgctgaggaagaagaagcgtatgagcgtatgtgctatctggtccaaatctggcag 120
Db 185 ctctgtgtgctgaggaagaagaagcgtatgagcgtatgtgctatctggtccaaatctggcag 244
QY 121 aaggaagacttcgaggaagcagcaggaaggaagcactgaaactgtggggcacaaggtgttc 180
Db 245 aaggaagacttcgaggaagcagcaggaaggaagcactgaaactgtggggcacaaggtgttc 304
QY 181 atgagagatgcaagcaggaaggttctgaggaaggttcaatctggcgccatccagtcacagc 240
Db 305 atgagagatgcaagcaggaaggttctgaggaaggttcaatctggcgccatccagtcacagc 364
QY 241 gcaactgtatgaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 300
Db 365 gcaactgtatgaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 424
QY 301 aaacaagctggaatcgccagcaggaaggaaggaaggaaggaaggaaggaaggaaggaag 360
Db 425 aaacaagctggaatcgccagcaggaaggaaggaaggaaggaaggaaggaaggaaggaag 484
QY 361 aaggggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 420
Db 485 aaggggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 544
QY 421 gtcatgctcccttgaggaagtgctgacatctcaaaacccgttcaaggccgacgaatgacctg 480
Db 545 gtcatgctcccttgaggaagtgctgacatctcaaaacccgttcaaggccgacgaatgacctg 604
QY 481 atgagatcgcaaaagcagcaggaattcccaatcccggttcaatcccaaggaacccgtggagc 540
Db 605 atgagatcgcaaaagcagcaggaattcccaatcccggttcaatcccaaggaacccgtggagc 664
QY 541 atgagatcgcaaaagcagcaggaattcccaatcccggttcaatcccaaggaacccgtggagc 600
Db 665 atgagatcgcaaaagcagcaggaattcccaatcccggttcaatcccaaggaacccgtggagc 724
QY 601 caagcgcctccagagctctcaaaagcagcaggaagcagcaggaagcagcaggaagcagcag 660
Db 725 caagcgcctccagagctctcaaaagcagcaggaagcagcaggaagcagcaggaagcagcag 784
QY 661 gacattctcgagatcgagttcaaaaagggtccctgtgaggttgaggtgaggtgaggtgag 720
Db 785 gacattctcgagatcgagttcaaaaagggtccctgtgaggttgaggtgaggtgaggtgag 844
QY 721 ggcacacacacacacacacacacacacacacacacacacacacacacacacacacac 780
Db 845 ggcacacacacacacacacacacacacacacacacacacacacacacacacacacac 844
QY 781 catggtctgggctgatttgacatctgagagacgccttcaattgaggtgaggtgaggtgag 904
Db 905 catggtctgggctgatttgacatctgagagacgccttcaattgaggtgaggtgaggtgag 964
QY 841 atctacagaagcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 900
Db 965 atctacagaagcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1024
QY 901 accatggagccgggaagtgagcaaaatcaaaagaagccctggggttgaatttgctgagctg 960
Db 1025 accatggagccgggaagtgagcaaaatcaaaagaagccctggggttgaatttgctgagctg 1084
QY 961 gtgtacacgggttctgagacagcccttgagtgaggttgcctgcacatgacgtgacgaag 1020

```

```

RESULT 4
US-09-652-126-9702
: Sequence 9702, Application US/09652126
: GENERAL INFORMATION:
: APPLICANT: Shvjan, Andrew W.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: FILE REFERENCE: 1600,1185-001
: CURRENT APPLICATION NUMBER: US/09/652,126
: PRIOR FILING DATE: 2000-08-30
: PRIOR APPLICATION NUMBER: 60/151,132
: NUMBER OF SEQ ID NOS: 10051
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9702
: LENGTH: 1656
: TYPE: DNA
: ORGANISM: Homo sapiens

```


Db 1085 gtgtatccgttctctgcaagccctgagtgtaatttgctccgcaactgcatcgccaag 1144
Qy 1021 tcccaagagcagatggaaggaagtgcaagtgctgctccctcaaggccaggtgtatc 1080
Db 1145 tcccaagagcagatggaaggaagtgcaagtgctgctccctcaaggccaggtgtatc 1204
Qy 1081 ctgagcgaggtcccaactgctctctacatgaagagctgtgtacatgaagtgcaag 1140
Db 1205 ctgagcgaggtcccaactgctctctacatgaagagctgtgtacatgaagtgcaag 1264
Qy 1141 ggtgattatgagccaactgatactgcaaggtgtcaatcaatcaatccctcagctgaag 1200
Db 1265 ggtgattatgagccaactgatactgcaaggtgtcaatcaatcaatccctcagctgaag 1324
Qy 1201 gaatatcatgctctccagaggaagtgtaactgccaataag 1239
Db 1325 gaatatcatgctctccagaggaagtgtaactgccaataag 1363

RESULT 5
US-09-652-816-9327
; Sequence 9327, Application US/09652816
; GENERAL INFORMATION:
; APPLICANT: Gutierrez-Ramos, Jose-Carlos
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.1177-001
; CURRENT APPLICATION NUMBER: US/09/652.816
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152.111
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9647
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9327
; LENGTH: 1656
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-652-816-9327

Query Match 99.5%; Score 1232.6; DB 25; Length 1656;
Best Local Similarity 99.7%; Pred. No. 6.1e-294;
Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 atgtcagaagaagctccgtgtgtctgtgacctgagtgccgctgagacacctgtgcatc 60
Db 125 atgtcagaagaagctccgtgtgtctgtgacctgagtgccgctgagacacctgtgcatc 184
Qy 61 ctctgtgtgtctgaagaaaggaagctatgacgtcatctgtcctatctgtgccaacatgtgcag 120
Db 185 ctctgtgtgtctgaagaaaggaagctatgacgtcatctgtcctatctgtgccaacatgtgcag 244
Qy 121 aaggaagacttgaggaagccaggaaggaaggaaggaaggaaggaaggaaggaaggaag 180
Db 245 aaggaagacttgaggaagccaggaaggaaggaaggaaggaaggaaggaaggaaggaag 304
Qy 181 attgaagacttgaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 240
Db 305 attgaagacttgaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 364
Qy 241 gcaacttgatgaggaagcgtactctctgtggaacctctctgtgccaagccctgcagtcgcgcg 300
Db 365 gcaacttgatgaggaagcgtactctctgtggaacctctctgtgccaagccctgcagtcgcgcg 424
Qy 301 aagaagatggaatgccaagcggaagcggaaggaaggaaggaaggaaggaaggaaggaagga 360
Db 425 aagaagatggaatgccaagcggaagcggaaggaaggaaggaaggaaggaaggaaggaagga 484
Qy 361 aaggggaagcatcaggtccggtttgagctcagctgtactcactgtgccccagataaag 420
Db 485 aaggggaagcatcaggtccggtttgagctcagctgtactcactgtgccccagataaag 544
Qy 421 gtcatgtctccctgtggaagatgctgaattctacaacgggttcaagggccgcaatgtactg 480

Db 545 gtcatgtctccctgtggaagatgctgaattctacaacgggttcaagggccgcaatgtactg 604
Qy 481 atggaatgagcaagaagcaagaaggaattcccatcccggtactcccaagaagaccgtgagac 540
Db 605 atggaatgagcaagaagcaagaaggaattcccatcccggtactcccaagaagaccgtgagac 664
Qy 541 atggaatgagcaagaagcaagaaggaattcccatcccggtactcccaagaagaccgtgagac 600
Db 665 atggaatgagcaagaagcaagaaggaattcccatcccggtactcccaagaagaccgtgagac 724
Qy 601 caagcgctcccaaggtctctacaagaaggaaggaaggaaggaaggaaggaaggaaggaag 660
Db 725 caagcgctcccaaggtctctacaagaaggaaggaaggaaggaaggaaggaaggaaggaag 784
Qy 661 gacattctgagatcgagttcaaaaaggggtccctgtgaaggtgacccaagtcgaagat 720
Db 785 gacattctgagatcgagttcaaaaaggggtccctgtgaaggtgacccaagtcgaagat 844
Qy 721 ggaaccacccacagaagcctctctgagctctcaatgtaacctgaagatcgcggaag 780
Db 845 ggaaccacccacagaagcctctctgagctctcaatgtaacctgaagatcgcggaag 904
Qy 781 catggtgtggccggtatgtgacatcgtggaaggaaggaaggaaggaaggaaggaaggaag 840
Db 905 catggtgtggccggtatgtgacatcgtggaaggaaggaaggaaggaaggaaggaaggaag 964
Qy 841 atctagagagcccaaggaagcaccatccttaccatgtaactatgaacatcgagccctc 900
Db 965 atctagagagcccaaggaagcaccatccttaccatgtaactatgaacatcgagccctc 1024
Qy 901 accatgagccgggaagtgcgcaaatcaacaagaagcctgtggtgaattgtctgtgcgcg 960
Db 1025 accatgagccgggaagtgcgcaaatcaacaagaagcctgtggtgaattgtctgtgcgcg 1084
Qy 961 gtgtatccggtttacgagcctgagcctgagtgtaattgttcgccaactgacgcgaag 1020
Db 1085 gtgtatccggtttacgagcctgagcctgagtgtaattgttcgccaactgacgcgaag 1144
Qy 1021 tcccaagagcagatggaaggaaggaagtgcaagtgctccgtctcgaagggccaaggtgtacatc 1080
Db 1145 tcccaagagcagatggaaggaaggaagtgcaagtgctccgtctcgaagggccaaggtgtacatc 1204
Qy 1081 ctgagcgaggtcccaactgctctctacatgaagagctgtgtacatgaagtgcaag 1140
Db 1205 ctgagcgaggtcccaactgctctctacatgaagagctgtgtacatgaagtgcaag 1264
Qy 1141 ggtgattatgagccaactgatactgcaaggtgtcaatcaatcaatccctcagctgaag 1200
Db 1265 ggtgattatgagccaactgatactgcaaggtgtcaatcaatcaatccctcagctgaag 1324
Qy 1201 gaatatcatgctctccagaggaagtgtaactgccaataag 1239
Db 1325 gaatatcatgctctccagaggaagtgtaactgccaataag 1363

RESULT 6
US-09-698-010-14283
; Sequence 14283, Application US/09698010
; GENERAL INFORMATION:
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.2029-001
; CURRENT APPLICATION NUMBER: US/09/698.010
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: 60/162.358
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 15684
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14283
; LENGTH: 1656

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-698-010-14283

```

Query Match	99.5%	Score 1232.6;	DB 27;	Length 1656;
Best Local Similarity	99.7%;	Pred. No. 6.1e-294;		
Matches 1235;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0

[illegible]

QY	961	ggtataccgggtttagagccctagccctgagatgtggaattgtccgcactgtcatgcgcag	1020
Db	1085	gtgtataccgggtttctctgtgcacagccctgagatgtggaattgtccgcacactgtcatgcgcag <td>1144</td>	1144
QY	1021	tcccaaggagcagatgtgaaaggaaagtgtcaggtgtccgtctctcaaggccagagtgtatc <td>1080</td>	1080
Db	1145	tcccaaggagcagatgtgaaaggaaagtgtcaggtgtccgtctctcaaggccagagtgtatc <td>1204</td>	1204
QY	1081	ctcgcccgagagctccaccacgtctctctcaaatgagagagctgtgagcatgtgaactgcag <td>1140</td>	1140
Db	1205	ctcgcccgagagctccaccacgtctctctcaaatgagagagctgtgagcatgtgaactgcag <td>1264</td>	1264
QY	1141	ggtgtattatgagccaactgtatgcaccacgggttcatcaaatcaatttccctcaaggcttgag <td>1200</td>	1200
Db	1265	ggtgtattatgagccaactgtatgcaccacgggttcatcaaatcaatttccctcaaggcttgag <td>1324</td>	1324
QY	1201	gaatatcatgtctccagagccaagatgcactgtccaatatg <td>1239</td>	1239
Db	1325	gaatatcatgtctccagagccaagatgcactgtccaatatg <td>1363</td>	1363

RESULT 7

```

? Sequence:8607, Application US/09698012
? GENERAL INFORMATION:
? APPLICANT: Geating, David P.
? APPLICANT: Holzman, Douglas A.
? TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
? TITLE OF INVENTION: THEREFOR
? FILE REFERENCE: 1600,2002-001
? CURRENT APPLICATION NUMBER: US/09/698,012
? CURRENT FILING DATE: 2000-10-27
? PRIOR APPLICATION NUMBER: 60/162,166
? PRIOR FILING DATE: 1999-10-28
? NUMBER OF SEQ ID NOS: 9719
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO: 8607
? LENGTH: 1656
? TYPE: DNA
? ORGANISM: Homo sapiens
? IS-09-698-012-8607

```

Query Match	99.5%	Score 1232.6;	DB 27;	Length 1656;
Best Local Similarity	99.7%;	Pred. No. 6.1e-294;		
Matches 1235;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0;

[illegible]

```

Db 485 aagggaacgatacagctccggttgagctcagctcactcactgcccccaataag 544
Qy 421 gtaattgctccctggaagatgctcgaattctacaaacgggttcaaggccgaatgactg 480
Db 545 gtaattgctccctggaagatgctcgaattctacaaacgggttcaaggccgaatgactg 604
Qy 481 atgagtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 540
Db 605 atgagtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 664
Qy 541 atgagtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 600
Db 665 atgagtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 724
Qy 601 caagagcctccaggttctctacagaagaacccaggaaccccaagcccccaacccct 660
Db 725 caagagcctccaggttctctacagaagaacccaggaaccccaagcccccaacccct 784
Qy 661 gacattcctcagatcagatctcaaaaagggttccctgtgaaggtgacccaagctcaagat 720
Db 785 gacattcctcagatcagatctcaaaaagggttccctgtgaaggtgacccaagctcaagat 844
Qy 721 ggcacaccccaacgaacactcctcttgagctctctatgtaaccgtgaagatcgcgagcag 780
Db 845 ggcacaccccaacgaacactcctcttgagctctctatgtaaccgtgaagatcgcgagcag 904
Qy 781 catgagcgtgagcgttatgtaacgttgagagacccgttcattggaatgaatcccgaggt 840
Db 905 catgagcgtgagcgttatgtaacgttgagagacccgttcattggaatgaatcccgaggt 964
Qy 841 atctacgaaccccaacgaacactcctcttgagctctctatgtaaccgtgaagccttc 900
Db 965 atctacgaaccccaacgaacactcctcttgagctctctatgtaaccgtgaagccttc 1024
Qy 901 accatgagacccgaagatgctcgaacaaatcaacaagcctgaggttgaaattgctgagctg 960
Db 1025 accatgagacccgaagatgctcgaacaaatcaacaagcctgaggttgaaattgctgagctg 1084
Qy 961 ggtatcagcgttctacagcctcagcctcagctgagttgtaattgtccgcacatctgacccag 1020
Db 1085 ggtatcagcgttctcagcagccctcagctgagttgtaattgtccgcacatctgacccag 1144
Qy 1021 tcccaagagcagatggaaggaagatgcaagtgctcgttcccaagggccaggtgtacatc 1080
Db 1145 tcccaagagcagatggaaggaagatgcaagtgctcgttcccaagggccaggtgtacatc 1204
Qy 1081 ctcggcgggaggtcccaactgtctctctacaaatgagagcgtgagcaatgagctgagc 1140
Db 1205 ctcggcgggaggtcccaactgtctctctacaaatgagagcgtgagcaatgagctgagc 1264
Qy 1141 ggtatgtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 1200
Db 1265 ggtatgtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 1324
Qy 1201 gaatatcatcgtctccagaagaagatgcatctgccaatag 1239
Db 1325 gaatatcatcgtctccagaagaagatgcatctgccaatag 1363

```

RESULT 8

```

US-09-710-281-5043
; Sequence 5043, Application US/09710281
; GENERAL INFORMATION:
; APPLICANT: Hunter, John J.
; APPLICANT: Shyjan, Andrew W.
; APPLICANT: Stidal, Hilde
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THERAPOR
; FILE REFERENCE: 1600.2036-001
; CURRENT APPLICATION NUMBER: US/09/710,281
; CURRENT FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: 60/164,254
; PRIOR FILING DATE: 1999-11-09

```

```

; NUMBER OF SEQ ID NOS: 5803
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5043
; LENGTH: 1656
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-710-281-5043

```

```

Query Match          99.5%; Score 1232.6; DB 28; Length 1656;
Best Local Similarity 99.7%; Pred. No. 6,1e-294;
Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 atgtccagcaaaagctcgtgtgtctgctcgtacagctgagcgtgacacactgacatc 60
Db 125 atgtccagcaaaagctcgtgtgtctgctcgtacagctgagcgtgacacactgacatc 184
Qy 61 ctggtgtgtgtgaagaagaacagctatgacgttatgtctatctgccaactgtgacag 120
Db 185 ctggtgtgtgtgaagaagaacagctatgacgttatgtctatctgccaactgtgacag 244
Qy 121 aaggaagacttcagagcagagcaggaagaagcactgaagcttggggccaaaagtgttc 180
Db 245 aaggaagacttcagagcagagcaggaagaagcactgaagcttggggccaaaagtgttc 304
Qy 181 atgagatgtcagcagagagcttgtgagagatcattctgcccgcacatccagctcagc 240
Db 305 atgagatgtcagcagagagcttgtgagagatcattctgcccgcacatccagctcagc 364
Qy 241 gcaatgtatagaacccgtacactcctcctgtgagcactctcttgcagagcccttgacccgc 300
Db 365 gcaatgtatagaacccgtacactcctcctgtgagcactctcttgcagagcccttgacccgc 424
Qy 301 aaacaagtgaataatcgcagccagcggagagggccaagtatgtgtccacagcgcacagga 360
Db 425 aaacaagtgaataatcgcagccagcggagagggccaagtatgtgtccacagcgcacagga 484
Qy 361 aaggggaacgatacagctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
Db 485 aaggggaacgatacagctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 544
Qy 421 gtaattgctccctggaagatgctcgaattctacaaacgggttcaaggccgaatgactg 480
Db 545 gtaattgctccctggaagatgctcgaattctacaaacgggttcaaggccgaatgactg 604
Qy 481 atgagtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 540
Db 605 atgagtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 664
Qy 541 atgagtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 600
Db 665 atgagtaacgaacgaacgaacggaattcccaatcccggttcaatcccaacccgttgagc 724
Qy 601 caagagcctccaggttctctacagaagaacccaggaaccccaagcccccaacccct 660
Db 725 caagagcctccaggttctctacagaagaacccaggaaccccaagcccccaacccct 784
Qy 661 gacattcctcagatcagatctcaaaaagggttccctgtgaaggtgacccaagctcaagat 720
Db 785 gacattcctcagatcagatctcaaaaagggttccctgtgaaggtgacccaagctcaagat 844
Qy 721 ggcacaccccaacgaacactcctcttgagctctctatgtaaccgtgaagatcgcgagcag 780
Db 845 ggcacaccccaacgaacactcctcttgagctctctatgtaaccgtgaagatcgcgagcag 904
Qy 781 catgagcgtgagcgttatgtaacgttgagagacccgttcattggaatgaatcccgaggt 840
Db 905 catgagcgtgagcgttatgtaacgttgagagacccgttcattggaatgaatcccgaggt 964
Qy 841 atctacgaaccccaacgaacactcctcttgagctctctatgtaaccgtgaagatcgcgac 900
Db 965 atctacgaaccccaacgaacactcctcttgagctctctatgtaaccgtgaagatcgcgac 1024

```

QY 901 accatggaccggaagtgcgcaaatcaacaagacctggctgtgaatttctgtgactg 960
 Db 1025 accatggaccggaagtgcgcaaatcaacaagacctggctgtgaatttctgtgactg 1084
 QY 961 gctatcccggttaccggtaccggtgagttgtaatttctgcacacgtacgtgcag 1020
 Db 1085 gctatcccggttaccggtaccggtgagttgtaatttctgcacacgtacgtgcag 1144
 QY 1021 tcccaaggagcagtggaaggaagtgcaagtgtccgtctcaaggcgcaagtgtacac 1080
 Db 1145 tcccaaggagcagtggaaggaagtgcaagtgtccgtctcaaggcgcaagtgtacac 1204
 QY 1081 ctggcccggaagtcgccacctgtctctcaatagagagctgtgacatgacgtgcag 1140
 Db 1205 ctggcccggaagtcgccacctgtctctcaatagagagctgtgacatgacgtgcag 1264
 QY 1141 gctgatatgagcaactatgaccggtgtacatcaaatccctcctgcgtgcag 1200
 Db 1265 gctgatatgagcaactatgaccggtgtacatcaaatccctcctgcgtgcag 1324
 QY 1201 gaatatcatcgtctccagagcaagtgcatctgccaatag 1239
 Db 1325 gaatatcatcgtctccagagcaagtgcatctgccaatag 1363

RESULT 9

US-09-716-920-1274
 ; Sequence 1274, Application US/09716920
 ; GENERAL INFORMATION:
 ; APPLICANT: Lloyd, Clare M.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; FILE REFERENCE: 1600, 2054-001
 ; CURRENT APPLICATION NUMBER: US/09/716, 920
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: 60/166,507
 ; NUMBER OF SEQ ID NOS: 1999-11-19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1274
 ; LENGTH: 1656
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-716-920-1274

Query Match 99.5%, Score 1232.6; DB 28; Length 1656;
 Best Local Similarity 99.7%, Pred. No. 6,1e-294;
 Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtccagcaagctccgtgtctctgctaccagtgagcgctggaacctgtgacac 60
 Db 125 atgtccagcaagctccgtgtctctgctaccagtgagcgctggaacctgtgacac 184
 QY 61 ctctgttgcgtgaaggaacagctatgacttgcctctctgccaatattggccag 120
 Db 185 ctctgttgcgtgaaggaacagctatgacttgcctctctgccaatattggccag 244
 QY 121 aaggaagcttcgaggaagcaggaaggaaggaagcttggtggcgaaggtgttc 180
 Db 245 aaggaagcttcgaggaagcaggaaggaaggaagcttggtggcgaaggtgttc 304
 QY 181 attagagatgcagcaggaagttgtgaggaagtcatctgcgcgcgcacatccagc 240
 Db 305 attagagatgcagcaggaagttgtgaggaagtcatctgcgcgcgcacatccagc 364
 QY 241 gcactgataggaacgctaccctctctggaacctcttcgagcgccctcatgcccgc 300
 Db 365 gcactgataggaacgctaccctctctggaacctcttcgagcgccctcatgcccgc 424
 QY 301 aaaaatgtgaatcgc 360
 Db 425 aaaaatgtgaatcgc 484

QY 361 aaggggaacgatacaggtctcgtttgagctcagctgtactactgtgcccccaataag 420
 Db 485 aaggggaacgatacaggtctcgtttgagctcagctgtactactgtgcccccaataag 544
 QY 421 gctatctctccctggaagatccttgaattctcaacacggttcaaggccgcaatgactg 480
 Db 545 gctatctctccctggaagatccttgaattctcaacacggttcaaggccgcaatgactg 604
 QY 481 atggagtcgcgaagaacacacggaattcccatcccggtcactcccaagacccgtggagc 540
 Db 605 atggagtcgcgaagaacacacggaattcccatcccggtcactcccaagacccgtggagc 664
 QY 541 atggatgaacatctatgacatagctacagagctgtgaattcctgggaaccccaaac 600
 Db 665 atggatgaacatctatgacatagctacagagctgtgaattcctgggaaccccaaac 724
 QY 601 caagcgctcccaagttcttaccagaaagcccaagaccgaagccccaacacccct 660
 Db 725 caagcgctcccaagttcttaccagaaagcccaagaccgaagccccaacacccct 784
 QY 661 gacattctgagatcaggttcaaaaaggggtccctgtgaagggtgacaaagtcagat 720
 Db 785 gacattctgagatcaggttcaaaaaggggtccctgtgaagggtgacaaagtcagat 844
 QY 721 ggcac 780
 Db 845 ggcac 904
 QY 781 catggtctgggctgatttgacatcgttgagagacgcttcaattggaatgaagtcagat 840
 Db 905 catggtctgggctgatttgacatcgttgagagacgcttcaattggaatgaagtcagat 964
 QY 841 atctacagagccacagagagacacacacacacacacacacacacacacacacacac 900
 Db 965 atctacagagccacagagagacacacacacacacacacacacacacacacacacac 1024
 QY 901 accatggaccggaagtgcgcaaatcaacaagacctggctgtgaatttctgtgactg 960
 Db 1025 accatggaccggaagtgcgcaaatcaacaagacctggctgtgaatttctgtgactg 1084
 QY 961 gctatcccggttaccggtaccggtgagttgtaatttctgcacacgtacgtgcag 1020
 Db 1085 gctatcccggttaccggtaccggtgagttgtaatttctgcacacgtacgtgcag 1144
 QY 1021 tcccaaggagcagtggaaggaagtgcaagtgtccgtctcaaggcgcaagtgtacac 1080
 Db 1145 tcccaaggagcagtggaaggaagtgcaagtgtccgtctcaaggcgcaagtgtacac 1204
 QY 1081 ctggcccggaagtcgccacctgtctctcaatagagagctgtgacatgacgtgcag 1140
 Db 1205 ctggcccggaagtcgccacctgtctctcaatagagagctgtgacatgacgtgcag 1264
 QY 1141 gctgatatgagcaactatgaccggtgtacatcaaatccctcctgcgtgcag 1200
 Db 1265 gctgatatgagcaactatgaccggtgtacatcaaatccctcctgcgtgcag 1324
 QY 1201 gaatatcatcgtctccagagcaagtgcatctgccaatag 1239
 Db 1325 gaatatcatcgtctccagagcaagtgcatctgccaatag 1363

RESULT 10

US-09-726-175-2581
 ; Sequence 2581, Application US/09726175
 ; GENERAL INFORMATION:
 ; APPLICANT: Geating, David P.
 ; APPLICANT: Fraser, Christopher C.
 ; APPLICANT: Donovan, Michael J.
 ; APPLICANT: Holtzman, Douglas A.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; FILE REFERENCE: 1600, 2054-001

Db	365	gacacgttatgagagaccgctaactcctctggaacactctcttcgacaggccctgatacgccgc	424
QY	301	aaacaaatgtagaaatcgcccaagcgaggaggggccaagtatgtgtccacgagcgccacagga	360
Db	425	aaacaaatgtagaaatcgcccaagcgaggaggggccaagtatgtgtccacgagcgccacagga	484
QY	361	aaaggggaacgatacaggtccgggttttgagtcaagctgctactaactatgccccagataaag	420
Db	485	aaaggggaacgatacaggtccgggttttgagtcaagctgctactaactatgccccagataaag	544
QY	421	gtcatgtccctcttgaggagatgtccttaactctcaaacccggttccaaggcgccaatgtacgtg	480
Db	545	gtcatgtccctcttgaggagatgtccttaactctcaaacccggttccaaggcgccaatgtacgtg	604
QY	481	atgtagatacgcaagaacaaacgaggaattcccaatcccggttcaactcccaagaaacccgttgagc	540
Db	605	atgtagatacgcaagaacaaacgaggaattcccaatcccggttcaactcccaagaaacccgttgagc	664
QY	541	atgtagatgaacaaactatgacatacagcttaacagagctgtgaaatcctctgtagaaccccaagaac	600
Db	665	atgtagatgaacaaactatgacatacagcttaacagagctgtgaaatcctctgtagaaccccaagaac	724
QY	601	caagcgctccacaggtgctcttaacagaaagacccaagaccagaacccaagcccccaacacccct	660
Db	725	caagcgctccacaggtgctcttaacagaaagacccaagaccagaacccaagcccccaacacccct	784
QY	661	gacatcttcgagatcagatcttaaaaaaagggtccctgtgtagaaggtagaacaagtltcaagat	720
Db	785	gacatcttcgagatcagatcttaaaaaaagggtccctgtgtagaaggtagaacaagtltcaagat	844
QY	721	ggcacaccacccaacagactcctcttgagactcttcatgttaacttgaaagcaagtctgcgggcaag	780
Db	845	ggcacaccacccaacagactcctcttgagactcttcatgttaacttgaaagcaagtctgcgggcaag	904
QY	781	catggcgttgagcgttatatgacaatctgtgagaaacgcgtctcaatgtgaaatgtagcccgaggt	840
Db	905	catggcgttgagcgttatatgacaatctgtgagaaacgcgtctcaatgtgaaatgtagcccgaggt	964
QY	841	atctcagagagaccccaagagacacacatccttaccatgtctcaattatagacatacgagccttc	900
Db	965	atctcagagagaccccaagagacacacatccttaccatgtctcaattatagacatacgagccttc	1024
QY	901	accatggaacccggagatgagcagcaaatcaaaacaaagagcgctgggttgaattgtctgtagctg	960
Db	1025	accatggaacccggagatgagcagcaaatcaaaacaaagagcgctgggttgaattgtctgtagctg	1084
QY	961	gtgtataccggttthaoggcctlaagcctctgagatgtgaattgtgtccgcaactgcatcgccag	1020
Db	1085	gtgtataccggttthaoggcctctctgcaacagcctctgagatgtgaattgtgtccgcaactgcatcgccag	1144
QY	1021	tcccaagagagagatltgaaagaggaagatgacagatggtccgtctccaagagggccaggtgtacac	1080
Db	1145	tcccaagagagagatltgaaagaggaagatgacagatggtccgtctccaagagggccaggtgtacac	1204
QY	1081	ctcgcccgaggagatcccaacagtctctctcaacatgtgagagcgctgtgagatataaagctgacg	1140
Db	1205	ctcgcccgaggagatcccaacagtctctctcaacatgtgagagcgctgtgagatataaagctgacg	1264
QY	1141	gtgtgatatgagccaactgatacgcaacgaggttcatcaacatcaatctccctcagctgag	1200
Db	1265	gtgtgatatgagccaactgatacgcaacgaggttcatcaacatcaatctccctcagctgag	1324
QY	1201	gaatatcatcgtctccagagagagttcaactgtgccaatag	1239
Db	1325	gaatatcatcgtctccagagagagttcaactgtgccaatag	1363

```

? TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
?
? TITLE OF INVENTION: THEREFOR
?
? FILE REFERENCE: 1600.2021-001
?
? CURRENT APPLICATION NUMBER: US/09/726,790
?
? CURRENT FILING DATE: 2000-11-30
?
? PRIOR APPLICATION NUMBER: 60/166,038
?
? PRIOR FILING DATE: 1999-11-30
?
? NUMBER OF SEQ ID NOS: 2814
?
? SOFTWARE: FastSeq for Windows Version 4.0
?
? SEQ ID NO 2021
?
? LENGTH: 1656
?
? TYPE: DNA
?
? ORGANISM: Homo sapiens
?
? US-09-726-790-2021

```

Query Match	99.5%	Score 1232.6;	DB 29;	Length 1656;
Best Local Similarity	99.7%;	Pred. No. 6.1e-294;		
Matches 1235;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0;

[illegible]

```

Db      |||||||
905      catggcgtgggcccgtatgacatcgtggaagacccgttcacatggaatggaagtcctccgaggt 964
QY      |||||||
841      atctacgagacccacgagcagccaccccttaccatgctacattagacatcgagccctc 900
Db      |||||||
965      atctacgagacccacgagcagccaccccttaccatgctacattagacatcgagccctc 1024
QY      |||||||
901      accatgagacccggaagctgcgcaaaatcaacaagagcctgggcttgaatttgcctgagctg 960
Db      |||||||
1025     accatgagacccggaagctgcgcaaaatcaacaagagcctgggcttgaatttgcctgagctg 1084
QY      |||||||
961      ggtgtaccggtttaccggtccctgagctggtgtaatttgcctgagcctgacatgcacag 1020
Db      |||||||
1085     ggtgtaccggtttctctgacacagccctgagctggtgtaatttgcctgagcctgacatgcacag 1144
QY      |||||||
1021     tcccgagagcagagtggaaggaagtgacagctgctcctcctcaaggccaggtgtacatc 1080
Db      |||||||
1145     tcccgagagcagagtggaaggaagtgacagctgctcctcctcaaggccaggtgtacatc 1204
QY      |||||||
1081     ctccgagagaggtcccaactgtctctctacatgagagctggtgagcagtgagctgacag 1140
Db      |||||||
1205     ctccgagagaggtcccaactgtctctctacatgagagctggtgagcagtgagctgacag 1264
QY      |||||||
1141     ggtgtatgagagcagctatgacacagcgggttcacatcaatcaatccctcagctgagag 1200
Db      |||||||
1265     ggtgtatgagagcagctatgacacagcgggttcacatcaatcaatccctcagctgagag 1324
QY      |||||||
1201     gaatacatcgtctccagagcaagtgactgcacaaatag 1239
Db      |||||||
1325     gaatacatcgtctccagagcaagtgactgcacaaatag 1363

```

RESULT 13

```

US-09-732-630-4186
; Sequence 4186, Application US/09732630
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Deeds, James
; APPLICANT: Berger, Allison
; APPLICANT: Zhao, Xunel
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF CERVICAL CANCER
; FILE REFERENCE: MRI-008B
; CURRENT APPLICATION NUMBER: US/09/732,630
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/169,681
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/171,350
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: US 60/189,315
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: US 60/203,791
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: US 60/210,600
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: US 60/220,114
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 4660
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 4186
; LENGTH: 1725
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1,2,1725
; OTHER INFORMATION: n = a,c,g, or t
US-09-732-630-4186

```

Query Match 99.5%; Score 1232.6; DB 29; Length 1725;
Best Local Similarity 99.7%; Pred. No. 6,1e-294;

```

Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY      |||||||
1      atgtccagcaagagctccgtgtgtctggtcctacagctggcgccctgacacccctgtgacac 60
Db      |||||||
132     atgtccagcaagagctccgtgtgtctggtcctacagctggcgccctgacacccctgtgacac 191
QY      |||||||
61      ctctgtgtgttgaaagaaacaaagctatgacgttactgtcctactgtgccaatctgtgcaag 120
Db      |||||||
192     ctctgtgtgttgaaagaaacaaagctatgacgttactgtcctactgtgccaatctgtgcaag 251
QY      |||||||
121     aagaaagactctgagagagcagcaagaaagacactgaagcttggcgccaagagtgctc 180
Db      |||||||
252     aagaaagactctgagagagcagcaagaaagacactgaagcttggcgccaagagtgctc 311
QY      |||||||
181     attagagatgcaagcagagaggttctgtgagagttcactctgtgcccgcacatccagtcacg 240
Db      |||||||
312     attagagatgcaagcagagaggttctgtgagagttcactctgtgcccgcacatccagtcacg 371
QY      |||||||
241     gcactgtatgagagcagcctacactcctgtggcaccctctctgccaagccctgacatgcgcgc 300
Db      |||||||
372     gcactgtatgagagcagcctacactcctgtggcaccctctctgccaagccctgacatgcgcgc 431
QY      |||||||
301     aacaagatgagaaatccgcagcggagagggccaagtatgtctccacgycgcacagga 360
Db      |||||||
432     aacaagatgagaaatccgcagcggagagggccaagtatgtctccacgycgcacagga 491
QY      |||||||
361     aagggaaacgatacagctcgttctgagctacagctgactacactcgtgcccccaagataag 420
Db      |||||||
442     aagggaaacgatacagctcgttctgagctacagctgactacactcgtgcccccaagataag 551
QY      |||||||
421     gtcaattgtcctctgagagatgtcctgaattctacacacgggttcacagggcgccgaatgactg 480
Db      |||||||
552     gtcaattgtcctctgagagatgtcctgaattctacacacgggttcacagggcgccgaatgactg 611
QY      |||||||
481     atggaatgagcaagcaacacaggaattccatcccggtactcccaagaacccgtgagac 540
Db      |||||||
612     atggaatgagcaagcaacacaggaattccatcccggtactcccaagaacccgtgagac 671
QY      |||||||
541     atggaatgagcaacccatgacacatacagctacagagctgtgaattcctgtgaaaccccaagac 600
Db      |||||||
672     atggaatgagcaacccatgacacatacagctacagagctgtgaattcctgtgaaaccccaagac 731
QY      |||||||
601     caagcgccctcagagctctacacagagacccagccagcccaagccccaacacccct 660
Db      |||||||
732     caagcgccctcagagctctacacagagacccagccagcccaagccccaacacccct 791
QY      |||||||
661     gaacatctcgaatcagagttcaaaaaaggggtccctgtgaaagtgtaaccaagtcaagat 720
Db      |||||||
792     gaacatctcgaatcagagttcaaaaaaggggtccctgtgaaagtgtaaccaagtcaagat 851
QY      |||||||
721     ggcacacacccacacagacccctctgtgagctcttcacatgacttgaacgaatctgcggcag 780
Db      |||||||
852     ggcacacacccacacagacccctctgtgagctcttcacatgacttgaacgaatctgcggcag 911
QY      |||||||
781     catgagctgagccgtatgtacatcgtgagaaacccgttcaatgtgaatgaaagtcgccaggt 840
Db      |||||||
912     catgagctgagccgtatgtacatcgtgagaaacccgttcaatgtgaatgaaagtcgccaggt 971
QY      |||||||
841     atctacgagacccacagcagcaccatcccttaccatgctcattttagacatcgagccctc 900
Db      |||||||
972     atctacgagacccacagcagcaccatcccttaccatgctcattttagacatcgagccctc 1031
QY      |||||||
901     accatgagacccggaagctgcgcaaaatcaacaagagcctgggcttgaatttgcctgagctg 960
Db      |||||||
1032     accatgagacccggaagctgcgcaaaatcaacaagagcctgggcttgaatttgcctgagctg 1091
QY      |||||||
961     ggtgtatcagcttaccggtcagccctgagctggaatttgcctgacacgtacatgcagcaag 1020
Db      |||||||
1092     ggtgtatcagcttaccggtcagccctgagctggaatttgcctgacacgtacatgcagcaag 1151
QY      |||||||
1021     tcccgagagagagtggaaggaagtgacagctgctcctcctcaaggccaggtgtacatc 1080
Db      |||||||
1152     tcccgagagagagtggaaggaagtgacagctgctcctcctcaaggccaggtgtacatc 1211

```


QY 1081 ctgagcgagagctccactgtctctctacatgagagctgtgagcattgaacgtgcag 1140
 Db 1212 ctgagcgagagctccactgtctctctacatgagagctgtgagcattgaacgtgcag 1271
 QY 1141 ggtgattatgagcaactgtatgcccaggttcatcaacaatccctcagctgaag 1200
 Db 1272 ggtgattatgagcaactgtatgcccaggttcatcaacaatccctcagctgaag 1331
 QY 1201 gaatacatcgtctccagagcaagctcactgccaatag 1239
 Db 1332 gaatacatcgtctccagagcaagctcactgccaatag 1370

RESULT 14 US-09-770-173-2339

Sequence 2339, Application US/09770173
 GENERAL INFORMATION:
 APPLICANT: White, David
 APPLICANT: Pan, Yang
 TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 TITLE OF INVENTION: THEREFOR
 FILE REFERENCE: 1600.2059-001
 CURRENT APPLICATION NUMBER: US/09/770.173
 CURRENT FILING DATE: 2001-01-26
 PRIOR APPLICATION NUMBER: US 60/178,876
 PRIOR FILING DATE: 2000-01-28
 NUMBER OF SEQ ID NOS: 3167
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2339
 LENGTH: 1755
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-770-173-2339

Query Match 99.5%; Score 1232.6; DB 30; Length 1755;
 Best Local Similarity 99.7%; Pred. No. 6.2e-294;
 Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtccagagcaagctcgtgtgtctgtgctcactgagctgagcgcttgacacccgtgcatc 60
 Db 131 atgtccagagcaagctcgtgtgtctgtgctcactgagctgagcgcttgacacccgtgcatc 190
 QY 61 ctgctgtgctgaagagcaagctgacgtacatgacatctgagcagcattgagcag 120
 Db 191 ctgctgtgctgaagagcaagctgacgtacatgacatctgagcagcattgagcag 250
 QY 121 aaggaagacttcagagagcaggaagagcactgaagcttggggccaataaagtgttc 180
 Db 251 aaggaagacttcagagagcaggaagagcactgaagcttggggccaataaagtgttc 310
 QY 181 atcgagatgtccagcagaggtttgtgagagatcatctcgccgagccatccagtcacag 240
 Db 311 atcgagatgtccagcagaggtttgtgagagatcatctcgccgagccatccagtcacag 370
 QY 241 gacactatgagagcagcactcctctggaacccctctctgagcagcgctgcatcgccgc 300
 Db 371 gacactatgagagcagcactcctctggaacccctctctgagcagcgctgcatcgccgc 430
 QY 301 aacgaagtgaatgcagcagcgagggggcgaagtatgtgtccacgagcgccacagga 360
 Db 431 aacgaagtgaatgcagcagcgagggggcgaagtatgtgtccacgagcgccacagga 490
 QY 361 aagggagagcagatccagcttgcagctcagcagctcactcactgagcccccaagtaag 420
 Db 491 aagggagagcagatccagcttgcagctcagcagctcactcactgagcccccaagtaag 550
 QY 421 gtcatgtctccttgagagatgctcctgaattctacaaccggttcaagggcgcaataactg 480
 Db 551 gtcatgtctccttgagagatgctcctgaattctacaaccggttcaagggcgcaataactg 610
 QY 481 atggagtgacgcaaaagcaacagggatcccatcccggtcactcccaagaacccgtlgagc 540

Db 611 atggagtgacgcaaaagcaacagggatcccatcccggtcactcccaagaacccgtlgagc 670
 QY 541 atggaatgaacacctatgcacatcagctagagctggaatccctgaggaaccccaagac 600
 Db 671 atggaatgaacacctatgcacatcagctagagctggaatccctgaggaaccccaagac 730
 QY 601 caagcgctccaggttcttacaagaagaccagagaccagccaaagcccccaacacct 660
 Db 731 caagcgctccaggttcttacaagaagaccagagaccagccaaagcccccaacacct 790
 QY 661 gacattctcgagatccaggttcaaaaaaggggtccctgtgaagtgacaaagtcagagat 720
 Db 791 gacattctcgagatccaggttcaaaaaaggggtccctgtgaagtgacaaagtcagagat 850
 QY 721 ggcacacccacacagactcctctgagctctcactgagcttgaagtgagagtgagggcag 780
 Db 851 ggcacacccacacagactcctctgagctctcactgagcttgaagtgagagtgagggcag 910
 QY 781 catggctgggctgattgacatcgttgagagaccgcttcaatggaaatgagtcaggagt 840
 Db 911 catggctgggctgattgacatcgttgagagaccgcttcaatggaaatgagtcaggagt 970
 QY 841 atcagagagccacagcagcagcaccatcccttaccatgctcattgagacatgagccttc 900
 Db 971 atcagagagccacagcagcagcaccatcccttaccatgctcattgagacatgagccttc 1030
 QY 901 accatgagccgggaagtgcgcaaaatcaaaacagagccttgagcttgaatttctgagctg 960
 Db 1031 accatgagccgggaagtgcgcaaaatcaaaacagagccttgagcttgaatttctgagctg 1090
 QY 961 gtgtacacgggtttagcagcagctgagctgagtgatgtgtgtgtgtgtgtgtgtgtgtgt 1020
 Db 1091 gtgtacacgggtttagcagcagctgagctgagtgatgtgtgtgtgtgtgtgtgtgtgtgt 1150
 QY 1021 tcccaagagcagatggaagaggaagtgcagctgctcctcctcaagggcgagctgacac 1080
 Db 1151 tcccaagagcagatggaagaggaagtgcagctgctcctcctcaagggcgagctgacac 1210
 QY 1081 ctgagcgagagctcccaactgtctctctacatgagagctgtgagcatgaacgtgcag 1140
 Db 1211 ctgagcgagagctcccaactgtctctctacatgagagctgtgagcatgaacgtgcag 1270
 QY 1141 ggtgattatgagcaactgtatgcccaggttcatcaacaatccctcagctgaag 1200
 Db 1271 ggtgattatgagcaactgtatgcccaggttcatcaacaatccctcagctgaag 1330
 QY 1201 gaatacatcgtctccagagcaagctcactgccaatag 1239
 Db 1331 gaatacatcgtctccagagcaagctcactgccaatag 1369

RESULT 15 US-09-760-475-983

Sequence 983, Application US/09760475
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 FILE REFERENCE: PT249
 CURRENT APPLICATION NUMBER: US/09/760.475
 PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
 NUMBER OF SEQ ID NOS: 4122
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 983
 LENGTH: 2057
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (2036)
 OTHER INFORMATION: n equals a,t,g, or c
 NAME/KEY: SITE

LOCATION: (2038)
OTHER INFORMATION: n equals a,t,g, or c
US-09-760-475-983

Query Match 99.5%; Score 1232.6; DB 30; Length 2057;
Best Local Similarity 99.7%; Pred. No. 6,4e-294;
Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 atgtccagcaaaagctccgtgtgtctgtgctacagctggtggtccctggtgacactgtgtatc 60
Db 533 atgtccagcaaaagctccgtgtgtctgtgctacagctggtggtccctggtgacactgtgtatc 592
QY 61 ctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
Db 593 ctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 652
QY 121 aagaaagacttcgaggaagcaggaagaaaggaagcactgaagcttggggccaaaggtgttc 180
Db 653 aagaaagacttcgaggaagcaggaagaaaggaagcactgaagcttggggccaaaggtgttc 712
QY 181 attgagatgtcagcaggaaggttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240
Db 713 attgagatgtcagcaggaaggttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 772
QY 241 gcaactgtatgaggaagcgttacctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 300
Db 773 gcaactgtatgaggaagcgttacctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 832
QY 301 aaacaagtgtgaatcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 360
Db 833 aaacaagtgtgaatcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 892
QY 361 aaggggaacagatcagatcagatcagatcagatcagatcagatcagatcagatcagatcagatc 420
Db 893 aaggggaacagatcagatcagatcagatcagatcagatcagatcagatcagatcagatcagatc 952
QY 421 gtcatgtctccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 953 gtcatgtctccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1012
QY 481 atgtgagtaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaag 540
Db 1013 atgtgagtaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaag 1072
QY 541 atgtgagtaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaag 600
Db 1073 atgtgagtaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaagcaaaag 1132
QY 601 caagcgtctccaggtctctacacagaaagcagcagcagcagcagcagcagcagcagcagcagc 660
Db 1133 caagcgtctccaggtctctacacagaaagcagcagcagcagcagcagcagcagcagcagcagc 1192
QY 661 gacatctctgagatcagatcagatcagatcagatcagatcagatcagatcagatcagatcagatc 720
Db 1193 gacatctctgagatcagatcagatcagatcagatcagatcagatcagatcagatcagatcagatc 1252
QY 721 ggcacacccacacacacacacacacacacacacacacacacacacacacacacacacacacacac 780
Db 1253 ggcacacccacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1312
QY 781 catgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840
Db 1313 catgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1372
QY 841 atttacaagacccacacacacacacacacacacacacacacacacacacacacacacacacacacac 900
Db 1373 atttacaagacccacacacacacacacacacacacacacacacacacacacacacacacacacacac 1432
QY 901 accatgtaccggtggaagtgtgcgaataacaaacaaagcctgtgtgtgtgtgtgtgtgtgtgtgt 960
Db 1433 accatgtaccggtggaagtgtgcgaataacaaacaaagcctgtgtgtgtgtgtgtgtgtgtgtgt 1492

```

```

QY 961 gtgtataccggtttacagcctagccctgagtgtaattgttcggccactgtgacgtgcaag 1020
Db 1493 gtgtataccggtttacagcctagccctgagtgtaattgttcggccactgtgacgtgcaag 1552
QY 1021 tcccaagagcaggtgtgaaaggaaggtgacaggtgtccgtctcctcaagggcaggtgtacalc 1080
Db 1553 tcccaagagcaggtgtgaaaggaaggtgacaggtgtccgtctcctcaagggcaggtgtacalc 1612
QY 1081 ctccgagcaggtgtcccaactgtctctctacaaatgaggaagcgtgtgagcaatgaaagtgcaag 1140
Db 1613 ctccgagcaggtgtcccaactgtctctctacaaatgaggaagcgtgtgagcaatgaaagtgcaag 1672
QY 1141 ggtgtatataagcacaactatgtccacacaggtttcatcaatcaatcctcctcaggtcagag 1200
Db 1673 ggtgtatataagcacaactatgtccacacaggtttcatcaatcaatcctcctcaggtcagag 1732
QY 1201 gaatatcatgtctccagagcaaggtcactgtgccaatag 1239
Db 1733 gaatatcatgtctccagagcaaggtcactgtgccaatag 1771

```

Search completed: February 12, 2002, 14:27:58
Job time: 6091 sec

Wed Feb 13 07:36:04 2002

us-09-775-693-1.rmpm

Page 14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:10:07 ; Search time 4918.34 Seconds
(without alignments)
598.044 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239
Sequence: 1 atgtcagcaagagctcgcgt.....gcaagtcactgcacaaatag 1239

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 2043436 seqs, 118699438 residues

Total number of hits satisfying chosen parameters: 4086872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Pending_Patents_NA_New.*
1: /cgn2_6/p/ptodata/2/pna/pct_NEW_COMB.seq.*
2: /cgn2_6/p/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/p/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/p/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/p/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/p/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/p/ptodata/2/pna/US11_NEW_COMB.seq.*
8: /cgn2_6/p/ptodata/2/pna/US12_NEW_COMB.seq.*
9: /cgn2_6/p/ptodata/2/pna/US13_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1231	99.4	1571	US-09-981-353-86	Sequence 86, App1
2	1199.6	96.8	1828	US-09-898-888A-13594	Sequence 13594, A
3	1183.8	95.5	2004	US-60-340-187-216	Sequence 216, App
4	1183.8	95.5	2005	US-60-340-187-330	Sequence 330, App
5	742.8	60.0	1871	US-60-340-187-951	Sequence 951, App
6	554.6	44.8	987	US-09-898-888A-22919	Sequence 22919, A
7	503	40.6	503	US-09-998-598-467	Sequence 467, App
8	503	40.6	503	US-09-998-598-702	Sequence 702, App
9	503	40.6	503	US-09-998-598-907	Sequence 907, App
10	503	40.6	503	US-09-998-598-924	Sequence 924, App
11	499.8	40.3	503	US-09-998-598-1365	Sequence 1365, App
12	467.4	37.7	469	US-09-998-598-787	Sequence 787, App
13	457.4	36.9	479	US-09-933-524A-54674	Sequence 54674, A
14	454	36.6	538	US-10-029-386-8869	Sequence 8869, App
15	453.2	36.6	465	US-09-922-340-266	Sequence 266, App
16	453.2	36.6	465	US-09-922-340A-266	Sequence 266, App
17	434.8	35.1	499	US-09-998-598-883	Sequence 883, App
18	432.8	34.9	474	US-09-904-013-19540	Sequence 19540, A
19	432	34.9	452	US-09-933-524A-63203	Sequence 63203, A
20	430	34.7	443	US-09-933-524A-54673	Sequence 54673, A
21	422	34.1	454	US-09-898-888A-22065	Sequence 22065, A
22	421.8	34.0	453	US-09-939-397-38994	Sequence 38994, A
23	405.4	32.7	408	US-09-939-397-1772	Sequence 1772, App
24	404.8	32.7	492	US-10-029-386-22593	Sequence 22593, A

25	403	32.5	412	US-09-925-564-28045	Sequence 28045, A
26	398	32.1	398	US-09-927-875A-6305	Sequence 6305, A
27	397.8	32.1	426	US-09-933-524A-49321	Sequence 49321, A
28	389.8	31.5	435	US-09-933-524A-58820	Sequence 58820, A
29	385.8	31.1	400	US-09-925-564-6718	Sequence 6718, App
30	384.2	31.0	505	US-10-029-386-9059	Sequence 9059, App
31	383.2	30.9	385	US-09-939-397-1771	Sequence 16995, A
32	382.2	30.8	403	US-09-933-524A-92776	Sequence 92776, A
33	379.6	30.6	401	US-09-925-564-28304	Sequence 28304, A
34	376.6	30.4	404	US-09-939-397-47	Sequence 11846, A
35	376	30.3	377	US-09-975-640-11846	Sequence 11846, A
36	371.4	30.0	459	US-09-975-640A-11846	Sequence 28044, A
37	367.4	29.7	468	US-09-925-564-28044	Sequence 28044, A
38	367.4	29.7	468	US-09-925-564-28986	Sequence 21043, A
39	354.4	28.6	382	US-09-925-564-21043	Sequence 29730, A
40	353	28.5	394	US-09-904-013-29730	Sequence 29730, A
41	352.2	28.4	359	US-09-925-564-6780	Sequence 6780, App
42	351.2	28.3	385	US-09-975-640A-26735	Sequence 26735, A
43	350	28.2	351	US-09-975-640A-26735	Sequence 26735, A
44	350	28.2	351	US-09-939-397-32843	Sequence 32843, A
45	349.8	28.2	363	US-09-939-397-32843	Sequence 32843, A

ALIGNMENTS

RESULT 1
US-09-981-353-86
Sequence 86, Application US/09981353
GENERAL INFORMATION:
APPLICANT: Lasek, Amy W.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981,353
CURRENT FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 86
LENGTH: 1571
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No: 1543330CB1
US-09-981-353-86

Query Match	99.4%	Score 1231	DB 6	Length 1571
Best Local Similarity	99.6%	Pred. No. 0		
Matches 1234	Conservative	0	Mismatches	5
			Indels	0
			Gaps	0
QY 1	atgtcagcaagagctcgcgttctgtgctcctacatgtgcgagccttgcacatctgtgcatc	60		
DB 89	atgtcagcaagagctcgcgttctgtgctcctacatgtgcgagccttgcacatctgtgcatc	148		
QY 61	ctctgtgtgtgaggaagaaagcgtatgcgtcatctgtgctcctacatgtgcgagccttgcacatctgtgcatc	120		
DB 149	ctctgtgtgtgaggaagaaagcgtatgcgtcatctgtgctcctacatgtgcgagccttgcacatctgtgcatc	208		
QY 121	aaggaagacttgcaggaagcgtatgcgtcatctgtgctcctacatgtgcgagccttgcacatctgtgcatc	180		
DB 209	aaggaagacttgcaggaagcgtatgcgtcatctgtgctcctacatgtgcgagccttgcacatctgtgcatc	268		
QY 181	atgaagatgtcaggaagcgtatgcgtcatctgtgctcctacatgtgcgagccttgcacatctgtgcatc	240		
DB 269	atgaagatgtcaggaagcgtatgcgtcatctgtgctcctacatgtgcgagccttgcacatctgtgcatc	328		
QY 241	gcactgtatgaggaagcgtatgcgtcatctgtgctcctacatgtgcgagccttgcacatctgtgcatc	300		
DB 329	gcactgtatgaggaagcgtatgcgtcatctgtgctcctacatgtgcgagccttgcacatctgtgcatc	388		
QY 301	aaacagtggaatgcagcagcgtgaggggccaagtatgttccacagcgccacagga	360		

```

?
? CURRENT FILING DATE: 2001-07-30
? PRIOR APPLICATION NUMBER: US/09/340,623
? PRIOR FILING DATE: 1999-06-28
? PRIOR APPLICATION NUMBER: US 09/205,070
? PRIOR FILING DATE: 1998-12-03
? NUMBER OF SEQ ID NOS: 45207
? SOFTWARE: SeqSIO for Windows Version 3.0
? SEQ ID NO 13594
? LENGTH: 1828
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(1828)
? OTHER INFORMATION: n = A,T,C or G
?
US-09-838-868A-13594

```

2000

QY	1	atgtccagcaaaaggtctccgtgtgtcttgcgcctacagttgacggtccgtgcacactctgtacac	6
Db	76	atgtccagcaaaaggtctccgtgtgtcttgcgcctacagttgacggtccgtgcacactctgtacac	1
QY	61	ctctgtgtgtgtctgtgaagaatacaaggtactatgacgttcattgtccattctgtccaaattgtccag	1
Db	136	ctctgtgtgtgtctgtgaagaatacaaggtactatgacgttcattgtccattctgtccaaattgtccag	1
QY	121	aaggaagaactctcgtgaagaaccccgaggaagaagcatctgaagctttggtgccaagaagtgttc	1
Db	196	aaggaagaactctcgtgaagaaccccgaggaagaagcatctgaagctttggtgccaagaagtgttc	2
QY	181	attgagagatgtcaagcagggagtttgtgtgaagagttcaactctgcgcgcacatccagccacgc	2
Db	256	attgagagagatgtcaagcagggagtttgtgtgaagagttcaactctgcgcgcacatccagccacgc	3
QY	241	gcaactgtatgaaagacgcgtctacactctctgtggaactctctgtgca -ggccctgtacgcgcg	2
Db	316	gcaactgtatgaaagacgcgtctacactctctgtggaactctctgtgcaaggtccctgtacatgcgcg	3
QY	300	caaaacaagtgaggaatactgcgcgcacagcgaggagggccaaatgtatgtctccac -ggcgccacag	3
Db	376	caaaacaagtgaggaatactgcgcgcacagcgaggagggccaaatgtatgtctccac -ggcgccacag	4
QY	359	gaaaggggaagacatcagttccgcgttttgagctcaagctgtctacactctgtgccccagataa	4
Db	436	gaaaggggaagacatcagttccgcgttttgagctcaagctgtctacactctgtgccccagataa	4
QY	419	aggtcatgtgcctctgtgaagatgtcctgtaattctthacaacccgtttcaaaagggcgcgaatgac	4
Db	496	aggtcatgtgcctctgtgaagatgtcctgtaattctthacaacccgtttcaaaagggcgcgaatgac	5
QY	479	tgatgtgagttac -gcaaaagcacaacccggtattcccatcccggttcattcccaagaaccgtgtg	5
Db	556	tgatgtgagttacgtaaggaagaacacacggtattcccatcccggttcattcccaagaaccgtgtg	6
QY	538	agcatgatatagaactcctatgtacaatcagctactagaggtcttgaaactctgtgaagaaccacag	5
Db	616	agcatgatatagaactcctatgtacaatcagctactagaggtcttgaaactctgtgaagaaccacag	6
QY	588	aaccaaagcgtctcaagttctctacacggaagcccaaggaacccaagcccaaacccccaaccc	6
Db	676	aaccaaagcgtctcaagttctctacacggaagcccaaggaacccaagcccaaacccccaaccc	7
QY	658	ccgtgaacattctcgtagatcgtagtttcaaaaaaggggtccctgtgaagggtgacacaagttcaag	7
Db	736	ccgtgaacattctcgtagatcgtagtttcaaaaaaggggtccctgtgaagggtgacacaagttcaag	8
QY	718	gatgtgacacaaccaacagaacctctcttgagacttcttaactgtacactgaacgaagctcggtgc	7
Db	796	gatgtgacacaaccaacagaacctctctcttgagacttcttaactgtacactgaacgaagctcggtgc	8

```

Oy 778 aagatgacgtgagcgtatgacatcgtgagaaacccgttcatlgtgaatgaatccga 837
    |||||||
Db 856 aagatgacgtgagcgtatgacatcgtgagaaacccgttcatlgtgaatgaatccga 915
Oy 838 ggtatctagagacccagcagcagcaccatccttaccatgtctattagaacatcagagcc 897
    |||||||
Db 916 ggtatctagagacccagcagcagcaccatccttaccatgtctattagaacatcagagcc 975
Oy 898 ttaccatgagccggaagtgcgcaaatcaaaagagcctgggcttgaatctgtcga 957
    |||||||
Db 976 ttaccatgagccggaagtgcgcaaatcaaaagagcctgggcttgaatctgtcga 1035
Oy 958 ctgtgtatctacgggttaccggttagccctgagtggtgaatttgcgcgcctgacgccc 1017
    |||||||
Db 1036 ctgtgtatctacgggttaccggttagccctgagtggtgaatttgcgcgcctgacgccc 1095
Oy 1018 aagtcacagagcaggtggaaggaagtgcaagtgctcgtcccaagggcagtgctac 1077
    |||||||
Db 1096 aagtcacagagcaggtggaaggaagtgcaagtgctcgtcccaagggcagtgctac 1155
Oy 1078 atcctcgccgagagtcctccactgtctctcaaatgagagcgtgtgagcagtgacgtg 1137
    |||||||
Db 1156 atcctcgccgagagtcctccactgtctctcaaatgagagcgtgtgagcagtgacgtg 1215
Oy 1138 caggtgtattatgagcgaactatgacacgaggttcccaaatcaatccctcagagctg 1197
    |||||||
Db 1216 caggtgtattatgagcgaactatgacacgaggttcccaaatcaatccctcagagctg 1275
Oy 1198 aaggaatatcatcgtctccagagcaggttcaatgccaatag 1239
    |||||||
Db 1276 aaggaatatcatcgtctccagagcaggttcaatgccaatag 1317

```

RESULT 3

US-60-340-187-216

Sequence 216, Application US/60340187

GENERAL INFORMATION:

```

APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Jian-Rui
APPLICANT: Ghosh, Malabika
APPLICANT: Meng, Gezh
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
                Polypeptides
FILE REFERENCE: 813
CURRENT APPLICATION NUMBER: US/60/340,187
CURRENT FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: US 09/577,409
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: PCT/US01/04927
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 09/519,705

```

```

PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/574,454
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: PCT/US01/04941
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: US 09/649,167
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US01/08631
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/552,929
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 09/770,160
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: PCT/US01/08656
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 09/577,408
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: PCT/US01/14827
PRIOR FILING DATE: 2001-05-16
NUMBER OF SEQ ID NOS: 1192
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO 216
LENGTH: 2004
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (95)...(346)
US-60-340-187-216

```

```

Query Match          95.5%; Score 1183.8; DB 9; Length 2004;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1188; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 atgtccagcaagagctccgtgttctgtcctacagtgagcggccttgacactcgtgac 60
    |||||||
Db 481 atgtccagcaagagctccgtgttctgtcctacagtgagcggccttgacactcgtgac 540
Oy 61 ctggtgtgtcgaaggaacaaagctatgacgtcattgtcctatctgccaatgtggcag 120
    |||||||
Db 541 ctggtgtgtcgaaggaacaaagctatgacgtcattgtcctatctgccaatgtggcag 600
Oy 121 aaggaagacttcgaaggaagcgaaggaaggaaggaagcctgaagcttggggccaaaagtgctc 180
    |||||||
Db 601 aaggaagacttcgaaggaagcgaaggaaggaaggaagcctgaagcttggggccaaaagtgctc 660
Oy 181 attgaagatcgaaggaaggttctgtgaagagttcatcttgccggcgaatccagtcagc 240
    |||||||
Db 661 attgaagatcgaaggaaggttctgtgaagagttcatcttgccggcgaatccagtcagc 720
Oy 241 gcaatgtatgagacgcgtactcctcctggtgacactctcttggcaggccctgacatgcggc 300
    |||||||
Db 721 gcaatgtatgagacgcgtactcctcctggtgacactctcttggcaggccctgacatgcggc 780
Oy 301 aaacaagtggaatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
    |||||||
Db 781 aaacaagtggaatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Oy 361 aaggggaacgatacagtgctcgtttagagtcagctgactctcactgagcccccagataaag 420
    |||||||
Db 841 aaggggaacgatacagtgctcgtttagagtcagctgactctcactgagcccccagataaag 900
Oy 421 gcaatgtcctcctgagagatgctcctgaattctaaacccggttcaagggccgaatgacgtg 480
    |||||||
Db 901 gcaatgtcctcctgagagatgctcctgaattctaaacccggttcaagggccgaatgacgtg 960
Oy 481 atggaagtaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 540
    |||||||
Db 961 atggaagtaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 1020

```

```

? PRIOR APPLICATION NUMBER: US 09/560, 875
? PRIOR FILING DATE: 2000-04-27
? PRIOR APPLICATION NUMBER: PCT/US01/03800
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 09/515, 126
? PRIOR FILING DATE: 2000-02-28
? PRIOR APPLICATION NUMBER: US 09/577, 409
? PRIOR FILING DATE: 2000-05-18
? PRIOR APPLICATION NUMBER: PCT/US01/04927
? PRIOR FILING DATE: 2001-02-26
? PRIOR APPLICATION NUMBER: US 09/519, 705
? PRIOR FILING DATE: 2000-03-07
? PRIOR APPLICATION NUMBER: US 09/574, 454
? PRIOR FILING DATE: 2000-05-19
? PRIOR APPLICATION NUMBER: PCT/US01/04941
? PRIOR FILING DATE: 2001-03-05
? PRIOR APPLICATION NUMBER: US 09/540, 217
? PRIOR FILING DATE: 2000-03-31
? PRIOR APPLICATION NUMBER: US 09/649, 167
? PRIOR FILING DATE: 2000-08-23
? PRIOR APPLICATION NUMBER: PCT/US01/08631
? PRIOR FILING DATE: 2001-03-30
? PRIOR APPLICATION NUMBER: US 09/552, 929
? PRIOR FILING DATE: 2000-04-18
? PRIOR APPLICATION NUMBER: US 09/770, 160
? PRIOR FILING DATE: 2001-01-26
? PRIOR APPLICATION NUMBER: PCT/US01/08656
? PRIOR FILING DATE: 2001-04-18
? PRIOR APPLICATION NUMBER: US 09/577, 408
? PRIOR FILING DATE: 2000-05-18
? PRIOR APPLICATION NUMBER: PCT/US01/14827
? NUMBER OF SEQ ID NOS: 1192
? SOFTWARE: PC_FL-Genes Version 6.0
? SEQ ID NO 330
? LENGTH: 2005
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (254)..(2005)
US-60-340-187-330

Query Match          95.5%; Score 1183.8; DB 9; Length 2005;
Best Local Similarity 99.4%; Pctd. No. 0;
Matches 1188; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 atgtccagcaaaaggtccgtgtgtctgtgcttcagtcacagtgcgcgccttgagcaactcgtgcatc 60
Db 482 atgtccagcaaaaggtccgtgtgtctgtgcttcagtcacagtgcgcgccttgagcaactcgtgcatc 541
QY 61 ctccgtgtgcgcgaaggaagaaggcctctagccgtcatcttgccctatctcggccaatgtggcag 120
Db 542 ctccgtgtgcgcgaaggaagaaggcctctagccgtcatcttgccctatctcggccaatgtggcag 601
QY 121 aaggaagaactcgaaggaagccaggaagaaggaagcactgaagccttggggcccaaaaagtgcttc 180
Db 602 aaggaagaactcgaaggaagccaggaagaaggaaggaagcactgaagccttggggcccaaaaagtgcttc 661
QY 181 attagaagattcagccagggagagtttctgtggagagattcatcctctgcccgcacatccagtcacagc 240
Db 662 attagaagattcagccagggagagtttctgtggagagattcatcctctgcccgcacatccagtcacagc 721
QY 241 gcaactgtataggaacgcgtactactcctctgggacactctcttgcagaagccctgcacatgcgcgc 300
Db 722 gcaactgtataggaacgcgtactactcctctgggacactctcttgcagaagccctgcacatgcgcgc 781
QY 301 aaaaagaatgtgaatctgcccagcgggaaggggccaaglatgtgtccacaagcgccacaagga 360
Db 782 aaaaagaatgtgaatctgcccagcgggaaggggccaaglatgtgtccacaagcgccacaagga 841
QY 361 aagggagaacatcaggtctcgtgttgtagtcaagcgtcactaccatctgccccccacagataaag 420

```



```

|||||
Db 181 attgagatgctcagcaggaggttggagaggtatcatctgcccgcattccagtcacg 240
QY 241 gacgtatagagaccgctactcctccggacac-tctcttgcaggccctgcatcgccg 299
Db 241 gacgtatagagaccgctactcctccggacac-tctcttgcaggccctgcatcgccg 300
QY 300 caacaagtgtgaaa-tcgccacgagaggggccaagtgtgtcccccgcgcacag 358
Db 301 caacaagtgtgaaattcgcgcacgagaggggccaagtgtgtcccccgcgcacag 360
QY 359 gaagagggaacatcagctcgggtttgagctcagctgtactactctg-ccccacagata 417
Db 361 gaagagggaacatcagctcgggtttgagctcagctgtactactctgccccacagata 420
QY 418 aagttatgtccctccggaggtgtcctgtaattctacaacgggttcaaggccgcgaatgac 477
Db 421 aagttatgtccctccggaggtgtcctgtaattctacaacgggttcaaggccgcgaatgac 480
QY 478 ctgattgagtaacgcaaaagc-aacacgggattcccatcccggtcactcccaagaccggtg 536
Db 481 ctgattgagtaacgcaaaagc-aacacgggattcccatcccggtcactcccaagaccggtg 540
QY 537 gagcatgtgtgaaacccctcacaatgacatcagctcgaagctggaatccctggaaccccaa 596
Db 541 gagcatgtgtgaaacccctcacaatgacatcagctcgaagctggaatccctggaaccccaa 600
QY 597 gaaccaagcgtctcaggtgtctctacacgaaagacccagaccagcccaagcccaaacac 656
Db 601 gaaccaagcgtctcaggtgtctctacacgaaagacccagaccagcccaagcccaaacac 660
QY 657 cccctgacattctcgagatcag-tcaaaaagggtccctgtg--aaggtgacacacgt 713
Db 661 cccctgacattctcgagatcaggtttcaaaaagggtccctgtgaaagtgagacacacgt 720
QY 714 --caagggtgacacacacacacac-cctcttgagctc-ttcatgtactgtgaagaaag 769
Db 721 tcaaggtatgagacacacacacacac-cctcttgagctc-ttcatgtactgtgaagaaag 780
QY 770 tcgcgagcaagcatggtcgctgggct--gtatgacatcgt-ggaagaccgtctcat--g 823
Db 781 tcgcgagcaagcatggtcgctgggct--gtatgacatcgt-ggaagaccgtctcat--g 840
QY 824 gaatgaagtcctcagaggtatctc--acgagaccacacagacacatcctttaccatgct 879
Db 841 aatgaagtcctcagaggtatctc-ttaccgagacccacagcagacacatcctttaccatgct 900
QY 880 cattagaacatcagagccttaccatgtgacggagagtgccgaatc--aaacaagc 936
Db 901 cattagaacatcagagccttaccatgtgacggagagtgccgaatc--aaacaagc 960
QY 937 ctgggcttgaattgtctgagc-tggtgtataccggtttac--ggcctagcctgaggtgt 993
Db 961 ctgggcttgaattgtgtgagctgtgtataccggttttcttgcacagccctgaggtgt 1020
QY 994 gaattgtcgcgcacatgcat--cgcaagctccagagcagagtggaaggg-aaagtgtca 1049
Db 1021 gaattgtcgcgcacatgcat--cgcaagctccagagcagagtggaaggg-aaagtgtca 1080
QY 1050 ggtgtcgtcctc--aaaggccaggtgtatacc--tgcggccggaggtcccccacgtgc 1103
Db 1081 ggtgtcgtcctc--aaaggccaggtgtatacc--tgcggccggaggtcccccacgtgc 1140
QY 1104 tctctacaatgtgagctgtgtgacatgaaagtgcaaggtgatatatga 1151
Db 1141 tctctacaatgtgagctgtgtgacatgaaagtgcaaggtgatatatga 1188

```

RESULT 6
 US-09-898-888A-22919
 ; Sequence 22919, Application US/09898888A
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.

```

; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; LIBRARIES
; FILE REFERENCE: 20411-748CON1
; CURRENT APPLICATION NUMBER: US/09/898, 888A
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/340, 623
; PRIOR FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: US 09/205, 070
; PRIOR FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 45207
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 22919
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(987)
; OTHER INFORMATION: n = A,T,C or G
US-09-898-888A-22919

```

Query Match 44.8%; Score 554.6; DB 6; Length 987;
 Best Local Similarity 87.4%; Pred. No. 2,6e-142;
 Matches 670; Conservative 0; Mismatches 4; Indels 93; Gaps 2;

```

QY 473 atgacctgtatgtgagtaacgaaacacacggaattccatcccggtcactcccaagacc 532
Db 2 atgacctgtatgtgagtaacgaaacacacggaattccatcccggtcactcccaagacc 61
QY 533 cgttgagcatgtatgtgaaacatctatgcatacgtatgaggtgtgaatccgtggagacc 592
Db 62 cgttgagcatgtatgtgaaacatctatgcatacgtatgaggtgtgaatccgtggagacc 121
QY 593 ccaagaacaaagcgtcctcaggtctctacacgaaagacccaggaacccaaagcccca 652
Db 122 ccaagaacaaagcgtcctcaggtctctacacgaaagacccaggaacccaaagcccca 181
QY 653 acacccctgacatctcgtgagatcgtatgtcaaaaagggtccctgtgaagtgaccacg 712
Db 182 acacccctgacatctc----- 198
QY 713 tcaagatgtgacacacacacacacacacacacacacacacacacacacacacacacacac 772
Db 199 ----ggatgtgacacacacacacacacacacacacacacacacacacacacacacacacac 254
QY 773 cgggcaagcatgtggtcggtatgtgacatcgttggagaaacgcgtctcatgtgaatgaa 832
Db 255 cgggcaagcatgtggtcggtatgtgacatcgttggagaaacgcgtctcatgtgaatgaa 314
QY 833 cccgaggtatctcagagacccacagcagacacacacacacacacacacacacacacacacac 892
Db 315 cccgaggtatctcagagacccacagcagacacacacacacacacacacacacacacacacac 374
QY 893 aggccttccatgtgacccggaggtgcycaaaatcaacaagcgtgtggaattgtg 952
Db 375 aggccttccatgtgacccgggaa----- 397
QY 953 ctgagctgtgtatacacgtttacgacctagccctgagtgtaattgttcgcgcacatgca 1012
Db 398 -----tgtatccggttctctgcaacagccctgagtgtaattgttcgcgcacatgca 448
QY 1013 tgcgaagtcccaagagcaggtgtgaaagggaaagtgtcgtctcctcaaggccagg 1072
Db 449 tgcgaagtcccaagagcaggtgtgaaagggaaagtgtcgtctcctcaaggccagg 508
QY 1073 tgtacatctcgcgcggaggtccacacgtctctctacaaatgagaggtgtgagatga 1132
Db 509 tgtacatctcgcgcggaggtccacacgtctctctacaaatgagaggtgtgagatga 568
QY 1133 acgtgcaaggtgtatataagcaactgtatgccacgggtgtatcaacaatcaatccctca 1192
Db 569 acgtgcaaggtgtatataagcaactgtatgccacgggtgtatcaacaatcaatccctca 628

```


QY 1193 ggcctgaagaatatactgcttcacagcaagctcaccataag 1239
 |||
 Db 629 ggcctgaagaatatactgcttcacagcaagctcaccataag 675

RESULT 7

US-09-998-598-467/c
 ; Sequence 467, Application US/09998598
 ; GENERAL INFORMATION:
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Chenault, Ruth A.
 ; APPLICANT: Meagher, Madelein Joy
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.561
 ; CURRENT APPLICATION NUMBER: US/09/998,598
 ; CURRENT FILING DATE: 2001-11-16
 ; NUMBER OF SEQ ID NOS: 2606
 ; SOFTWARE: Corixa Invention Disclosure Database
 ; SEQ ID NO 467
 ; LENGTH: 543
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-998-598-467

Query Match 43.7%; Score 541; DB 7; Length 543;
 Best Local Similarity 99.6%; Pred. No. 1,2e-138;
 Matches 541; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 393 ctgctactactgccccccagataaagctatgctccctgagagatgctgaattcta 452
 |||
 Db 543 ytgctactactgccccccagataaagctatgctccctgagagatgctgaattcta 484
 QY 453 caaccggttcaagggccgaatgactgagctacgcaaaacgaacggatctccat 512
 |||
 Db 483 caaccggttcaagggccgaatgactgagctacgcaaaacgaacggatctccat 424
 QY 513 ccggttactcccaagaacccggtgagctatgagaaacctatgacatacgaatcaga 572
 |||
 Db 423 cccggttactcccaagaacccggtgagctatgagaaacctatgacatacgaatcaga 364
 QY 573 ggcctgaatccttgagaaaccccaagaacgaagcctccagctctctacagaagacca 632
 |||
 Db 363 ggcctgaatccttgagaaaccccaagaacgaagcctccagctctctacagaagacca 304
 QY 633 ggaaccagccaaagcccccaaacacccctgacatctcgaagatcgaattcaaaaaaggggt 692
 |||
 Db 303 ggaaccagccaaagcccccaaacacccctgacatctcgaagatcgaattcaaaaaaggggt 244
 QY 693 cccgttgaaggtgacaaagctcaaggaatgacaccccaagaacccctccttgagctctt 752
 |||
 Db 243 cccgttgaaggtgacaaagctcaaggaatgacaccccaagaacccctccttgagctctt 184
 QY 753 catgtactgaagcaagctcgaggcaagcatgagcgttgagccgataatgacatcgttgagaa 812
 |||
 Db 183 catgtactgaagcaagctcgaggcaagcatgagcgttgagccgataatgacatcgttgagaa 124
 QY 813 ccgcttcaattggaatgaagctccgaaggtatctaagaagcccccaagcagccatccttta 872
 |||
 Db 123 ccgcttcaattggaatgaagctccgaaggtatctaagaagcccccaagcagccatccttta 64
 QY 873 caatgtcaattggaatgaagctcccaatgacagccggaggaagtgcgcaaatcaaca 932
 |||
 Db 63 caatgtcaattggaatgaagctcccaatgacagccggaggaagtgcgcaaatcaaca 4
 QY 933 agg 935
 |||
 Db 3 AGG 1

RESULT 8

US-09-998-598-702
 ; Sequence 702, Application US/09998598
 ; GENERAL INFORMATION:
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Chenault, Ruth A.
 ; APPLICANT: Meagher, Madelein Joy
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.561
 ; CURRENT APPLICATION NUMBER: US/09/998,598
 ; CURRENT FILING DATE: 2001-11-16
 ; NUMBER OF SEQ ID NOS: 2606
 ; SOFTWARE: Corixa Invention Disclosure Database
 ; SEQ ID NO 702
 ; LENGTH: 503
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-998-598-702

Query Match 40.6%; Score 503; DB 7; Length 503;
 Best Local Similarity 100.0%; Pred. No. 3,2e-128;
 Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 ctgctactactgccccccagataaagctatgctccctgagagatgctgaattcta 452
 |||
 Db 1 ctgctactactgccccccagataaagctatgctccctgagagatgctgaattcta 60
 QY 453 caaccggttcaagggccgaatgactgagctacgcaaaacgaacggatctccat 512
 |||
 Db 61 caaccggttcaagggccgaatgactgagctacgcaaaacgaacggatctccat 120
 QY 513 ccggttactcccaagaacccggtgagctatgagaaacctatgacatacgaatcaga 572
 |||
 Db 121 ccggttactcccaagaacccggtgagctatgagaaacctatgacatacgaatcaga 180
 QY 573 ggcctgaatccttgagaaaccccaagaacgaagcctccagctctctacagaagacca 632
 |||
 Db 181 ggcctgaatccttgagaaaccccaagaacgaagcctccagctctctacagaagacca 240
 QY 633 ggaaccagccaaagcccccaaacacccctgacatctcgaagatcgaattcaaaaaaggggt 692
 |||
 Db 241 ggaaccagccaaagcccccaaacacccctgacatctcgaagatcgaattcaaaaaaggggt 300
 QY 693 cccgttgaaggtgacaaagctcaaggaatgacaccccaagaacccctccttgagctctt 752
 |||
 Db 301 cccgttgaaggtgacaaagctcaaggaatgacaccccaagaacccctccttgagctctt 360
 QY 753 catgtactgaagcaagctcgaggcaagcatgagcgttgagccgataatgacatcgttgagaa 812
 |||
 Db 361 catgtactgaagcaagctcgaggcaagcatgagcgttgagccgataatgacatcgttgagaa 420
 QY 813 ccgcttcaattggaatgaagctccgaaggtatctaagaagcccccaagcagccatccttta 872
 |||
 Db 421 ccgcttcaattggaatgaagctccgaaggtatctaagaagcccccaagcagccatccttta 480
 QY 873 caatgtcaattggaatgaagctcccaatgacagccggaggaagtgcgcaaatcaaca 935
 |||
 Db 481 caatgtcaattggaatgaagctcccaatgacagccggaggaagtgcgcaaatcaaca 503

RESULT 9

US-09-998-598-907
 ; Sequence 907, Application US/09998598
 ; GENERAL INFORMATION:
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Chenault, Ruth A.
 ; APPLICANT: Meagher, Madelein Joy
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER

FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 907
LENGTH: 503
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-907

Query Match 40.6%; Score 503; DB 7; Length 503;
Best Local Similarity 100.0%; Pred. No. 3.2e-128;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 393 ctgtctactactgccccccagataaaggtcatgtctccctggagatgctgaattcta 452
DB 1 ctgtctactactgccccccagataaaggtcatgtctccctggagatgctgaattcta 60
QY 453 caaccggttcaaggccgcaatgacatgagtagcgaagcaacacgggattcccat 512
DB 61 caaccggttcaaggccgcaatgacatgagtagcgaagcaacacgggattcccat 120
QY 513 ccggttactctcccaagaaccggtgagtagatgagaacctctacatcagctacga 572
DB 121 ccggttactctcccaagaaccggtgagtagatgagaacctctacatcagctacga 180
QY 573 ggtctgaatcctctggaagaccacgaagaacgctctcccaagttctctacagaagacca 632
DB 181 ggtctgaatcctctggaagaccacgaagaacgctctcccaagttctctacagaagacca 240
QY 633 ggaaccagccaagcccccaacacccctgacatctctcgaagtcgaattcaaaaagggt 692
DB 241 ggaaccagccaagcccccaacacccctgacatctctcgaagtcgaattcaaaaagggt 300
QY 693 cccgtgaagtgagcaacgctcaagtagatgagcaacacccacgacctcttgagctctt 752
DB 301 cccgtgaagtgagcaacgctcaagtagatgagcaacacccacgacctcttgagctctt 360
QY 753 catgtacctgaagtagcgcgagcaagcatggtggtgagctgattacatcgtagagaa 812
DB 361 catgtacctgaagtagcgcgagcaagcatggtggtgagctgattacatcgtagagaa 420
QY 813 ccgcttcatggaatgagtagtcccgaggtatctacagagaccacgagcagaccatcttla 872
DB 421 ccgcttcatggaatgagtagtcccgaggtatctacagagaccacgagcagaccatcttla 480
QY 873 ccagtccatttagacatcgagg 895
DB 481 ccagtccatttagacatcgagg 503
```

RESULT 10
US-09-998-598-924
Sequence 924, Application US/09998598

GENERAL INFORMATION:

APPLICANT: Stoik, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 924
LENGTH: 503
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-924

Query Match 40.6%; Score 503; DB 7; Length 503;
Best Local Similarity 100.0%; Pred. No. 3.2e-128;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 393 ctgtctactactgccccccagataaaggtcatgtctccctggagatgctgaattcta 452
DB 1 ctgtctactactgccccccagataaaggtcatgtctccctggagatgctgaattcta 60
QY 453 caaccggttcaaggccgcaatgacatgagtagcgaagcaacacgggattcccat 512
DB 61 caaccggttcaaggccgcaatgacatgagtagcgaagcaacacgggattcccat 120
QY 513 ccggttactctcccaagaaccggtgagtagatgagaacctctacatcagctacga 572
DB 121 ccggttactctcccaagaaccggtgagtagatgagaacctctacatcagctacga 180
QY 573 ggtctgaatcctctggaagaccacgaagaacgctctcccaagttctctacagaagacca 632
DB 181 ggtctgaatcctctggaagaccacgaagaacgctctcccaagttctctacagaagacca 240
QY 633 ggaaccagccaagcccccaacacccctgacatctctcgaagtcgaattcaaaaagggt 692
DB 241 ggaaccagccaagcccccaacacccctgacatctctcgaagtcgaattcaaaaagggt 300
QY 693 cccgtgaagtgagcaacgctcaagtagatgagcaacacccacgacctcttgagctctt 752
DB 301 cccgtgaagtgagcaacgctcaagtagatgagcaacacccacgacctcttgagctctt 360
QY 753 catgtacctgaagtagcgcgagcaagcatggtggtgagctgattacatcgtagagaa 812
DB 361 catgtacctgaagtagcgcgagcaagcatggtggtgagctgattacatcgtagagaa 420
QY 813 ccgcttcatggaatgagtagtcccgaggtatctacagagaccacgagcagaccatcttla 872
DB 421 ccgcttcatggaatgagtagtcccgaggtatctacagagaccacgagcagaccatcttla 480
QY 873 ccagtccatttagacatcgagg 895
DB 481 ccagtccatttagacatcgagg 503
```

RESULT 11

US-09-998-598-1365/c
Sequence 1365, Application US/09998598

GENERAL INFORMATION:

APPLICANT: Stoik, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 1365
LENGTH: 503
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-1365

Query Match 40.3%; Score 499.8; DB 7; Length 503;
Best Local Similarity 99.6%; Pred. No. 2.4e-127;
Matches 501; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 393 ctgtctactactgccccccagataaaggtcatgtctccctggagatgctgaattcta 452
DB 503 CTGCTACTACTGCCCCCAGATTAAGTCAATTCCTCTGAGAGATGCTGAATTCCTA 444
```

```

QY 443 caaccggttcaaggccgcaatgacatgagatgacgcaagaacacggaattccat 512
DB 443 CAACCGGTTCAAGGCGCCATGACCTGATGAGATGACCAAGAACACGGATCCCAT 384
QY 513 cccggtacatcccaagaacccggtgagatgagatgagaacctaatgacatcagctaga 572
DB 383 CCCGGTCACTCCCAAGAACCGGTGAGCATGATGAGAACCTCATGCAATCACTAGCA 324
QY 573 ggcctggaatcctgaggaaccccaagaacacggtccctcagatctctacacgaagacca 632
DB 323 GGCTGGAATCTCGAGAACCCCAAGAACCAAGCCGCTCTCAGAGTCTTACAGAACCA 264
QY 633 ggaaccaagcaagaagcccaacaacccctgacatctcagatctcagatcgaagagagt 692
DB 263 GGAACCAAGCAAGCCCAACACCCCTGACATCTCTCAGATCTGAGTCAAAAAAGGGT 204
QY 693 cccctgtgaagtgacaaagtgcaagatgagccacccacacagacctctctggaacct 752
DB 203 CCTGTGAAGGTGACCAACGTCAGAGATGACACCAACCAACGACCTCTTGGAGCTTT 144
QY 753 catgtacacgaagtgacgagcgagcaagcatggtggtggtggtatgacatcgtgagaa 812
DB 143 CATGTACTGACGAGATGCGGACAGCATGCGTGGGCGCTATTGACATCTGTGAGAA 84
QY 813 cccgttcaatggaatgagtgccgaggtatctacgagaccccaagagcacatcctta 872
DB 83 CCGCTTCAATGGAATGAGTCCGAGGTATCTACGAGACCCCAAGACGACCATCTTTA 24
QY 873 ccatgtcattagacatcagag 895
DB 23 CCATGCTCATTTAGACATCGAG 1

```

```

RESULT 12
US-09-998-598-787
Sequence 787, Application US/09998598
GENERAL INFORMATION:

```

```

APPLICANT: Stolk, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 787
LENGTH: 469
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-787

```

```

Query Match 37.7%; Score 467.4; DB 7; Length 469;
Best Local Similarity 99.8%; Pred. No. 1.9e-118;
Matches 468; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 427 gctcctgagatgctgtaattctacaaacggttcaaggccgcaatgacatgagag 486
DB 1 gctcctgagatgctgtaattctacaaacggttcaaggccgcaatgacatgagag 60
QY 487 taagcaagaacacggaattccatcccggtcactcccaagaacccgtgagcatgat 546
DB 61 taagcaagaacacggaattccatcccggtcactcccaagaacccgtgagcatgat 120
QY 547 gagaacctatgacatcagctacgaggtgagatctctggaagaaccccaagaacgaag 606
DB 121 gagaacctatgacatcagctacgaggtgagatctctggaagaaccccaagaacgaag 180
QY 607 cctcaggtctacaaagaacccagagcccgcaagaaccccaacccctgacatc 666

```

```

DB 181 cctcaggtctctacacgaagaccagaccagcccaagcccccacacccctgacatt 240
QY 667 ctgagatcaggttcaaaaaagggttcctgtgaagtgacaaagtgacagtgagacc 726
DB 241 ctgagatcaggttcaaaaaagggttcctgtgaagtgacaaagtgagaccagtgagacc 300
QY 727 acccaacagacctcttgagatctctacatgacatgacgaagtgagcgggcaagcagtg 786
DB 301 acccaacagacctcttgagatctctacatgacatgacgaagtgagcgggcaagcagtg 360
QY 787 gtggcgcatatgacatctgtgagagacccgtctacatgagatgaagtcacaggtatctac 846
DB 361 gtggcgcatatgacatctgtgagagacccgtctacatgagatgaagtcacaggtatctac 420
QY 847 ggaaccacagcagcaccatccttaccatgcatgcatatgaacatcagag 895
DB 421 ggaaccacagcagcaccatccttaccatgcatgcatatgaacatcagag 469

```

```

RESULT 13
US-09-933-524A-54674
Sequence 54674, Application US/09933524A
GENERAL INFORMATION:

```

```

APPLICANT: Drmanac, Radoje T.
APPLICANT: Iabat, Ivan
APPLICANT: Stachne-Crain, Birgit
APPLICANT: Dickson, Mark
APPLICANT: Jones, Lee W.
TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
FILE REFERENCE: 774
CURRENT APPLICATION NUMBER: US/09/933,524A
CURRENT FILING DATE: 2001-08-20
PRIOR APPLICATION NUMBER: 09/528,409
PRIOR FILING DATE: 2000-05-17
NUMBER OF SEQ ID NOS: 116231
SOFTWARE: HY-Patent..pl Version 3.1
SEQ ID NO 54674
LENGTH: 479
TYPE: DNA
ORGANISM: Homo sapiens
US-09-933-524A-54674

```

```

Query Match 36.9%; Score 457.4; DB 6; Length 479;
Best Local Similarity 99.8%; Pred. No. 1.1e-115;
Matches 458; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 gtcacgaagaaggtcctggttctgtgctacagtgagcggtctgacacccctgacatcct 62
DB 20 gtcacgaagaaggtcctggttctgtgctacagtgagcggtctgacacccctgacatcct 79
QY 63 cgtgtggtcgaaggaacagcctatgacgtcatgctatctggtccacatctggcagaag 122
DB 80 cgtgtggtcgaaggaacagcctatgacgtcatgctatctggtccacatctggcagaag 139
QY 123 ggaagactcgaaggaacgaaggaaggaagcactgaagcttggtggccaaagaagtgttcat 182
DB 140 ggaagactcgaaggaacgaaggaaggaagcactgaagcttggtggccaaagaagtgttcat 199
QY 183 tgagatgtcgaaggaaggttggtgagagttacatctggtggccacacatccagtcacagc 242
DB 200 tgagatgtcgaaggaaggttggtgagagttacatctggtggccacacatccagtcacagc 259
QY 243 actgtatgagacgtctactcctctggtgacacacacacacacacacacacacacacacac 302
DB 260 actgtatgagacgtctactcctctggtgacacacacacacacacacacacacacacacac 319
QY 303 acaagtggaaatccgcaagcgagggggaaggaaggaatgtgtccacggcgcaaggaagaa 362
DB 320 acaagtggaaatccgcaagcgagggggaaggaaggaatgtgtgtccacggcgcaaggaagaa 379
QY 363 gggaagacatcaggtcgtgtgagctgactacatcagtcgacacacacacacacacacacac 422

```

Db 380 ggggaacgacatcagctccggttgagctcagctcgtactactacgcccccaataaagt 439
 Qy 423 catgtccctgagagatccctgtaattcacaacgggt 461
 Db 440 catgtccctgagagatcgtcgttaattcacaacgggt 478

RESULT 14

US-10-029-386-8869/C
 ; Sequence 8869, Application US/10029386
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: AEMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; NUMBER OF SEQ ID NOS: 3428
 ; SOFTWARE: Annomax Sequence Listing Engine Vers. 1.1
 ; SEQ ID NO 8869
 ; LENGTH: 538
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC005000.1
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.76
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
 ; OTHER INFORMATION: NE HIT: g114740527, EVALUATE 0.00e+00
 ; OTHER INFORMATION: SWISSPROT HIT: P00966, EVALUATE 3.00e-74
 ; OTHER INFORMATION: EST_HUMAN HIT: BE783168.1, EVALUATE 0.00e+00
 US-10-029-386-8869

Query Match

Best Local Similarity 36.6%; Score 454; DB 8; Length 538;
 Matches 492; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

Qy 104 tggccacatttggccagaagaagacttggagaagccagaagaagccactgaagctg 163
 Db 538 TGCCCAACACTGGCCAGAGGAAGACTTGAAGAGCCAGAGAGAGCCACTGAAGCTTG 479
 Qy 164 gggccaaaaggtgttcattgagatgtcagcagggagttgtgaggaagttcattggc 223
 Db 478 GGGGCAAAAAGGTGTTTCATTGAGGAAGTCAGCAAGAGTTGTGAAGAGTTTCATCTGGC 419
 Qy 224 cggccatcagtcagcgacgtgtagagacggtactctctgtggaacctcttgca 283
 Db 418 TGCCCATTCAGTCACGCGACTGTATGAGAGCACTACTCTCTGCACTCTCTCAACA 359
 Qy 284 ggcctcagtcagcccgcaaaaagtggaatcccgccagcgggaggggccaagtatgt 343
 Db 358 GGCCTCGATCCCGCAAAAACAGAGGAATCGCCAGCAGAGAGGAGGCAAGTAAGTGT 299
 Qy 344 ccacagcgccacgaagaagggaacgagtcagtcggttgaagctgaagttgtctacac 403
 Db 298 CCCACAGCGCCACAGAGGAAGGAATGATCAGTTCAGTTGAGCTTAACCTGCTACTCTC 239
 Qy 404 tgg-ccccccagataaagttatgtctccctcgtggaagttgctgaattctacaacgggtc 462
 Db 238 TGGCCCCCCAGATTAAGTATGCTCCCTGGAGATGCCCAAGTTCTACACAGGTTTC 179
 Qy 463 aagggcgcaatgaactgtatgagtaacgaagaacagggagttccatcccggtcact 522
 Db 178 AAGGTCCGAATGACCTTATGGAACACACAAGACACAGGAGTTCCTCCATCCAGTCACT 119
 Qy 523 cccagaacccgtgagcagatgagatgaacacctatgacatcagctcagaggtggaatc 582
 Db 118 CCCAAGAACCTGTGAGCAATGAGAGAACCTCATGATCAGATCAGCAATAGAGGTGGAATC 59

Qy 583 ctggaagaccccaagaacacagcgcctcagttctacacgaagaccagagaccag 640
 Db 58 TGGAGAACCTTAAGAACCAAGCATTTCCAGTCTCTCCAGAGAGACCCAGAGACCGG 1

RESULT 15

US-09-922-340-266
 ; Sequence 266, Application US/09922340
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
 ; FILE REFERENCE: 20411-726CON2
 ; CURRENT APPLICATION NUMBER: US/09/922,340
 ; PRIOR APPLICATION NUMBER: 2001-08-03
 ; PRIOR FILING DATE: 1999-07-14
 ; PRIOR APPLICATION NUMBER: US 09/004,182
 ; PRIOR FILING DATE: 1998-01-07
 ; NUMBER OF SEQ ID NOS: 12181
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 266
 ; LENGTH: 465
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(465)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-922-340-266

Query Match

Best Local Similarity 36.6%; Score 453.2; DB 6; Length 465;
 Matches 455; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 334 aagtaatgtccacagcgccacgaagaagggaagacagtcagtcggttgagctcagc 393
 Db 1 aagtaatgtccacagcgccacgaagaagggaagacagtcagtcggttgagctcagc 60
 Qy 394 tgcatactcgtgccccccagataaagttcattgtctccttgagagatgagctgaattac 453
 Db 61 tgcatactcgtgccccccagataaagttcattgtctccttgagagatgagctgaattac 120
 Qy 454 aaccggttaagggccgcgaatgacatgagtagtagtagtagtagtagtagtagtagtag 513
 Db 121 aaccggttaagggccgcgaatgacatgagtagtagtagtagtagtagtagtagtagtag 180
 Qy 514 ccggtacactcccaagaacccgttgagagcagtagtagtagtagtagtagtagtagtagtag 573
 Db 181 ccggtacactcccaagaacccgttgagagcagtagtagtagtagtagtagtagtagtagtag 240
 Qy 574 gctgtaactcgtgagagaaccccaagaacagcgctccaggtctctacacgaagaccag 633
 Db 241 gctgtaactcgtgagagaaccccaagaacagcgctccaggtctctacacgaagaccag 300
 Qy 634 gaccagcgaagaagcccaacaacccctgacatctcgtgagtagtagtagtagtagtagtag 693
 Db 301 gaccagcgaagaagcccaacaacccctgacatctcgtgagtagtagtagtagtagtagtag 360
 Qy 694 cctgtgaagtgagcaacagtcagtagtagtagtagtagtagtagtagtagtagtagtagtag 753
 Db 361 cctgtgaagtgagcaacagtcagtagtagtagtagtagtagtagtagtagtagtagtagtag 420
 Qy 754 atgtacactgaaaggaagtcgagggcgaagcagagcagtagtagtagtagtagtagtagtag 792
 Db 421 atgtacactgaaaggaagtcgagggcgaagcagtagtagtagtagtagtagtagtagtagtag 459

Wed Feb 13 07:36:05 2002

us-09-775-693-1.rnpn


```

117 yAlaThrGlyGlyAsnAspValAlaArgPheGluLeuSerCysTyr 134
401 CACTGGCCCCCGAATAAGATGATGCTCCCTGGAGATGGCTTAATTC 450
134 erLeuAlaProGlnIleLysValIleAlaProTyrArgMetProGluHe 150
451 TACACGGGTTCAGGGCCCAATGACCTGATGGAGTCCGAAAGACACA 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGluTyrAlaLysGlnI 167
501 CGGGATTCACATCCCGGTACTCCCAAGAACCCGTGAGCATGGATGAGA 550
167 sGlyLeuProIleProValThrProLysAsnProTyrSerMetAspGlu 184
551 ACCATGACATGACATGACGATGACGATGATGATGATGATGATGATG 600
184 snLeuMetHisIleSerTyrGluAlaGlyIleLeuGluAsnProLysAsn 200
601 CAGGCGCTCCAGGTCTCTACACAGAGACCCAGGACCCCAAGCCGCC 650
201 GlnAlaProProGlyLeuTyrThrLysThrGlnAspProAlaLysAlaPr 217
651 CACACCCCTGACATCTCCAGATCGATGCTCAAAAAGGGGCTCTGTA 700
217 oAsnThrProAspIleLeuGluIleGluPheLysGlyValProValI 234
701 AGGTGACCAACGTCAGATGACACCCAGACACCTCCCTGGAGCTC 750
234 yValAlThrAsnValLysArgGlyThrThrHisIleThrSerLeuGluLeu 250
751 TTCATGACCTGAACGACAGTGGGGGACAGATGCGGCGCTGATTA 800
251 PheMetLysLeuAsnGluValAlaGlyLysHisGlyValGlyArgIleAs 267
801 CATGTGAGAACCGCTTCATGATGATGATGATGATGATGATGATGATG 850
267 rIleValGluAsnArgPheIleGlyMetLysSerArgGlyIleTyrGluT 284
851 CCCGACGAGCACCATCTTACCATGCTCATTTAGACATCGAGCGCTTC 900
284 hProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAlaPhe 300
901 ACCATGACCGGAGAGTGGCAAAATCAACAAAGGCTGGGCTGAAAT 950
301 ThrMetAspArgGluValAlaArgLysIleLysGlnGlyLeuLysPhe 317
951 TGTGTGAGCGGTGTATACCGGTTAGCGCTAGCCCTAGTGAATTTG 1000
317 eAlaGluLeuValTyrThrGlyLeuArgProSerProGluCysGluPhe 334
1001 TCCGCGCATGCTCGCAAGTCCAGGAGGAGGAGGAGGAGGAGGAGG 1050
334 aArgHisCysIleLeuLysSerGlnGluArgValGluGlyLysValGln 350
1051 GTTCCGCTCCCAAGGCGCAGGTATACATCTCGCGCGGAGTCCCACT 1100
351 ValSerValLeuLysGlyGlnValTyrIleLeuGlyArgGluSerProLe 367
1101 GTCTCTACATGAGAGCTGTGAGCATGAGCATGAGCATGAGCATGAG 1150
367 uSerLeuTyrAsnGlnGluLeuValSerMetAsnValGlnGlyAspTyrG 384
1151 AGCCACTGATGACCGGGTTCATCAACATCAATCCCTCAGGCTGAAG 1200
384 LuProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATATCATGCTCCAGAGCATGACCTGCCAAA 1236
401 GluTyrHisArgLeuGlnSerLysValThrAlaLys 412
seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-775-693-7

```

```

seq_documentation_block:
; <sequence 7> Application 05/09/75693-1
; GENERAL INFORMATION:
; APPLICANT: Clark, Mike
; APPLICANT: Holtzberg, Frederick Wayne
; APPLICANT: Ensor, Charles Mark
; TITLE OF INVENTION: Methods for Predicting Susceptibility of Patients to Arginine
; TITLE OF INVENTION: Therapy
; FILE REFERENCE: PHOS-0060
; CURRENT APPLICATION NUMBER: US/09/775,693
; CURRENT FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-775-693-7

alignment_scores:
Quality: 2144.00 Length: 412
Ratio: 5.204 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-775-693-1 x US-09-775-693-7 ..
Align seq 1/1 to: US-09-775-693-7 from: 1 to: 412

1 ATGTCCAGCAAAAGGCTCCGTGTTCTGGCTACAGTGGCGGCTGACAC 50
1 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspTh 17
51 CTGTGATCCCTGCTGGCTGAGGACACAGGCTATGATGCTGCT 100
17 rSerCysIleLeuValTyrPheLysGlnGluIleGlyAspValIleAlaT 34
101 ATGTGGCCACATTTGGCCAGAAAGACTTCGAGAGAGCCAGAGAGAG 150
34 yLeuAlaAsnIleGlyGlnLysGluAspPheGluGluAlaArgLysLys 50
151 GCACGTAACCTGGGGCCCAAAAGGTTCATTGAGATGTCACAGAGA 200
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerArgL 67
201 GTTTGTGAGAGATTCATCTGGCGGCGCATCGATCCAGGCACTGTATG 250
67 uPheValGluGluPheIleTyrProAlaIleGlnSerSerAlaLeuTyrG 84
251 AGGACCGTACCTCTCTGGGACCTCTCTTGGCAGCGCTCATCGCCCG 300
84 LuAspArgTyrLeuLeuGlnThrSerLeuAlaArgProCysIleAlaArg 100
301 AAACAGTGAATCGCCCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 350
101 LysGlnValGluIleAlaGlnArgGlnGlyAlaLysTyrValSerHisG 117
351 CGCAGCAGAAAGGAGGACATGATGATGATGATGATGATGATGATGAT 400
117 yAlaThrGlyLysGlyAsnAspValAlaArgPheGluLeuSerCysTyrS 134
401 CACTGGCCCCCGAATAAGATGATGATGATGATGATGATGATGATGAT 450
134 erLeuAlaProGlnIleLysValIleAlaProTyrArgMetProGluPhe 150
451 TACACCGGTTCAGGGCCCAATGACCTGATGGAGTCCGAAAGACACA 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGluTyrAlaLysGlnI 167
501 CGGGATTCACATCCCGGTACTCCCAAGAACCCGTGAGCATGGATGAGA 550
167 sGlyLeuProIleProValThrProLysAsnProTyrSerMetAspGlu 184

```



```

551 ACCATGACATCAAGCTACGAGGCTGGATCTTGAGAACCCCAAGAAC 600
|||||
184 snleuethisileserlyrgluialaglylleugluasnprolyasn 200
|||||
601 CAAGGGCCCGAGGTCTCTACAGCAAGACCCGAGACCCCAAGAGCC 650
|||||
201 GlnAlaProProGlyLeuThrlystThrglnAspProAlaLysAlaPr 217
|||||
651 CAACACCCCTGACATCTCTGAGATCGAGTTCACAAAAGGGCTCCCTGGA 700
|||||
217 OasnthrProAspIleLeuGluIleGluPheLysGlyValProVal 234
|||||
701 AGGTACCAACGTCAGAGATGGCCACCACCACAGACCTCTTGAGAGCTC 750
|||||
234 yValThrAsnValLysAspGlyThrThnIsglnThrSerLeuGluLeu 250
|||||
751 TTCATGTATCTACAGAAAGTCCGGGCAAGCATGGCTGGCGGTATTGA 800
|||||
251 PheMetTyLLeuAsnGluValAlaGlyLysHisGlyValGlyArgIleAs 267
|||||
801 CATCTGTGAGAACCCCTCATCTGATGAAATGACCCGAGTATCTACAGA 850
|||||
267 rIleValGluAsnArgPheIleGlyMetLysSerArgGlyIleTyrgLut 284
|||||
851 CCCCAAGCAGGACCATCTTACATGCTCATTTAGACATCGAGGCGCTTC 900
|||||
284 hrProAlaGlyThrIleLeuTygHisAlaHisLeuAspIleGluAlaPhe 300
|||||
901 ACCATGACAGGGAAGTGGGCAAAATCAACAGAGCTGGCTGTAATAT 950
|||||
301 ThrMetAspArgGluValArgLysIleLysGlnGlyLeuGlyLeuLysSh 317
|||||
951 TCGTAGCGGTGTATACCGGTTTACGGCTAGCCCTGAGTGTGAATTGG 1000
|||||
317 eAlaGluLeuValTyThrGlyLeuArgProSerProGluCysGluPheV 334
|||||
1001 TCGGCCATGTCATCGCAAGTCCCAAGAGCGGAGGAGGAAAGTGCAG 1050
|||||
334 aAlaArgIscysIleAlaLysSerGlnGluArgValGluGlyLysValGln 350
|||||
1051 GTGTCCGTCTCAAGGGCCAGGCTATCATCTCTGGCCGGAGTCCCACT 1100
|||||
351 ValSerValLeuLysGlyGlnValTyThrLeuGlyArgGluSerProhe 367
|||||
1101 GTCCTCTCAATGAGAGCTGTGAGCATGAACGTGACGGGTGATTAG 1150
|||||
367 uSerLeuTygAsnGluGluLeuValSerMetLysValGlnGlyAspTyrg 384
|||||
1151 AGCCAACTGATGCAACGGGTTTCATCAACATCAATCCCTCAGGGTGAAG 1200
|||||
384 lUpThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
|||||
1201 GAATATCATGCTCTCCAGAGCAAGTCACTGCCAAA 1236
|||||
401 GluTygHisArgLeuGlnSerLysValThrAlaLys 412

```

seq_name: /cgn2_6/ptodata/2/paa/us60_COMB.pep:us-60-239-841-87

seq_documentation_block:

; Sequence 87, Application US/60239841

; GENERAL INFORMATION:

; APPLICANT: Lasek, Amy W.

; APPLICANT: Jones, David A.

; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER

; FILE REFERENCE: PA-0038 P

; CURRENT APPLICATION NUMBER: US/60/239, 841

; NUMBER OF SEQ ID NOS: 194

; SOFTWARE: PERL Program

; SEQ ID NO 87

; LENGTH: 412

; TYPE: PRT

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 1543330CD1
US-60-239-841-87

```

```

alignment_scores:
  Quality: 2123.00      Length: 412
  Ratio: 5.178          Gaps: 0
  Percent Similarity: 99.515  Percent Identity: 99.272

```

alignment_block:

US-09-775-693-1 x US-60-239-841-87 ..

Align seg 1/1 to: US-60-239-841-87 From: 1 to: 412

```

1 ATGTCCAGCAAGGCTCCGTGCTCTGGCTACAGTGGCGGCTGGACAC 50
|||||
1 MetSerSerLysGlySerValValLeuAlaTySerGlyLeuAspThr 17
|||||
51 CTGTCATCTCTCGTGTGCTGAGAAACAAGGCTATGAGTCACTTGCCT 100
|||||
17 rSerCysIleLeuValIrrPheLysGlnGlnGlyTyThrAspValIleAla 34
|||||
101 ATGTGGCCACATTTGGCCAGAGAGACTTTCAGAGAACCCAGAGAGAG 150
|||||
34 yLeuAlaAsnIleGlyGlnLysGluAspPheGlnGluAlaArgLysLys 50
|||||
151 GCATGAAAGCTTTGGGCGCAAAAGGTGTCTCATGAGATGTCCACAGAGA 200
|||||
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerAla 67
|||||
201 GTTGTGAGAGAGTTCATCTGCGCGCCATCCAGTCCAGCGCACTGTATG 250
|||||
67 uPheValGluLeuPheIleThrProAlaIleGlnSerSerAlaLeuTyrg 84
|||||
251 AGGACCGCTACCTCTGGGCACTCTCTTTCGACAGGCGCTGCATCGCCGC 300
|||||
84 lAspArgTyLeuLeuGlnGlyThrSerLeuAlaArgProCysIleAlaArg 100
|||||
301 AAACAAGTGAATTCGCCACAGGAGGAGGCGCAAGTATGTCTCCACAGG 350
|||||
101 LysGlnValGluIleAlaGlnArgGlnGlyAlaLysTyValSerHisGly 117
|||||
351 CGCCACAGGAAGGGAACGATCAGTCCGGTTGAGCTCAGCTGACTACT 400
|||||
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuSerCysTyrs 134
|||||
401 CACTGGCCCCCAGATAAAGGTCAATGCTCCCTGAGAGATGCCGTAATTC 450
|||||
134 eIleAlaIrrProGlnIleLysValIleAlaPrrThrArgMetProGluPhe 150
|||||
451 TACAAACGGTTCAGAGGCGCAATGACTGATGAGTACGCAAAAGCAACA 500
|||||
151 TygAsnArgPheLysGlyArgAsnAspLeuMetGluTygAlaLysGlnI 167
|||||
501 CGGATTCCTCCATCCCGGCTCACTCCAGAACCCGTGGAGCATGGATAGA 550
|||||
167 sGlyIleProIleProValIrrProLysAsnProIrrSerMetLyspAla 184
|||||
551 ACCTCATGACATCAGTACGAGGCTGGAATCTCGAGAAACCCCAAGAAC 600
|||||
184 snleuethisileserlyrgluialaglylleugluasnprolyasn 200
|||||
601 CAAGCGCTCCAGGTCTCTACAGAGAACCCAGAGACCCCAAGAGCC 650
|||||
201 GlnAlaProProGlyLeuTyThrLysThrGlnAspProAlaLysAlaPr 217
|||||
651 CAACACCCCTGACATCTCTGAGATCGAGTTCACAAAAGGGCTCCCTGGA 700
|||||
217 OasnthrProAspIleLeuGluIleGluPheLysGlyValProVal 234

```

```

701 AGGTGACCAACGTCAAGATGAGCAGCACCAGCACTCTTGAGATC 750
|||||
234 ysvaIThrAsnValIysAspGIyThrThrlsGlnThrSerLeuGIleu 250
|||||
751 TTATGTTACTGTAACAGATGCGGGGAGCAGCATGCGTGGCCGTATGA 800
|||||
251 PheMeTyrlLeuAsnGIleValAlaGIlyLysGIlyValGIlyArgIleAs 267
|||||
801 CATGTGAGAGACCGCTTCATTTGAATGAATGCCGAGGTATAGAGAG 850
|||||
267 pIleValGIleAsnThrPheIleGIleuTyLysSerArgGIlyIleTyGIleu 284
|||||
851 CCCGACGAGCAGCACCATTCTTACCATGCTTAAAGCATGAGGCGCTTC 900
|||||
284 hrProAlaGIlyThrlleuTyrlleuTyrlhisAlaHisleuAspIleGIlyAlaPhe 300
|||||
901 ACCATGAGACGGGAGAGTGGCCAAATCAACAAGCGCTGGCGTTGAAT 950
|||||
301 ThlMetAspArgGIlyValArgIlyLysGIlyLysGIlyLeuGIlyLeuLys 317
|||||
951 TGGTGAAGCTGTGATACGGTTTACGGCTAGCGCTGAGTGTGATTTG 1000
|||||
317 eAlaGIleuValIyThrlGIlyPheTrpHisSerProGIlyLysGIlyPhe 334
|||||
1001 TCCGCACTGCATCGCCAGATCCGAGAGCGAGTGAAGGAAAGTGCAG 1050
|||||
334 alArgHisCysIleAlaIyLysSerGIlyLysValArgGIlyLysValGIln 350
|||||
1051 GTGTCCGCTCTCAAGGGCCAGGTGTACATCTCGCGCGGAGTCCCACT 1100
|||||
351 ValSerAlaLeuLysGIlyLysValIyLleuGIlyArgGIlySerProIle 367
|||||
1101 GTCTCTTACATGAGAGAGCTGTGATGAGCATGAAGTGAAGGTGATTA 1150
|||||
367 uSerLeuTyrlAsnGIlyLysGIlyLeuValSerMetAsnValGIlnGIlyAspTyrG 384
|||||
1151 AGCCAACTGATGCCACCGGTTTCATCAACATTCCTCAGCGTGAAG 1200
|||||
384 IupProThrAspAlaThrlGIlyPheIleAsnIleAsnSerLeuArgLeuLys 400
|||||
1201 GAATATCATGTCTCCAGAGCAGGTTCACATCGCCAAA 1236
|||||
401 GlutylHisArgLeuGIlyLysValThrlAlaLys 412
|||||
seq_name: /cgn2_6/ptodata/2/paa/PCTus_COMB.pep:PCT-US00-05988-1594

seq_documentation_block:
; Sequence 1594, Application PC/TUS0005988
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Human Prostate Cancer Associated Gene Sequences and Polypeptides
; FILE REFERENCE: PAID1PCT
; CURRENT APPLICATION NUMBER: PCT/US00/05988
; CURRENT FILING DATE: 2000-03-08
; EARLIER APPLICATION NUMBER: 60/124,270
; EARLIER FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1594
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (22)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US00-05988-1594

```

alignment_scores:

Quality: 2123.00
Ratio: 5.178

Length: 412
Caps: 0

Percent Similarity: 99.515 Percent Identity: 99.272

alignment_block:
US-09-775-693-1 x PCT-US00-05988-1594 ..

Align seg 1/1 to: PCT-US00-05988-1594 from: 1 to: 442

```

1 ATGTCCAGCAAAAGCGCTCGGTGTGTGGCTACAGTGGCGGCGTGAAC 50
|||||
31 MetSerSerLysGIlySerValAlaLeuAlaTyrlSerGIlyLeuAspTh 47
|||||
51 CTCGTGCATCTCTGTGTGGCTGAAGAACAAAGCATGATGCTCATTTGCT 100
|||||
47 rSerCysIleLeuValTrpLeuLysGIlyLysGIlyLysArgIleAlaTr 64
|||||
101 ATCTGCGCAACATTTGGCCAGAAAGACACTTCGAGAAAGCCAGAGAAC 150
|||||
64 yrlLeuAlaAsnIleGIlyLysGIlyLysAspPheGIlyLysAlaArgLys 80
|||||
151 GCACGTGAAGCTTGGGCGCAAAAGGTGTTCATTTGAGATGTTCAGAGGA 200
|||||
81 AlaLeuLysLeuGIlyAlaLysLysValPheIleGIlyLysPheArgGI 97
|||||
201 GTTTGTGAGGAGTTCATCTGCGCGCCATCCAGTCCAGCGCACTGTATG 250
|||||
97 uPheValGIlyLysGIlyLysPheIleTrpProAlaIleGIlySerLysLeuTyrG 114
|||||
251 AGGACCGCTACTCTCTCGGCGACCTCTCTTCCAGAGCGCTGCATGCGCC 300
|||||
114 IuAspArgTyrLeuLeuGIlyThrSerLeuAlaArgProCysIleAlaArg 130
|||||
301 AAACAAGTGAAGATGCCCGAGCGGAGGGGSCAAGTATGTGTCCACGG 350
|||||
131 LysGIlyValGIlyIleAlaGIlyLysValAlaLysTyrValSerHisGI 147
|||||
351 CGCCACAGAAAGGGGAGACGATCAAGTCCGTTTGAAGTCAAGTCACTACT 400
|||||
147 yAlaThrlGIlyLysGIlyAsnAspGIlyValArgPheGIlyLysSerCysTyrS 164
|||||
401 CACTGCGCCCCCAGATTAAGTCAATGTCTCCCTGAGAGATCCCTGAATTC 450
|||||
164 erLeuAlaProGIlyLysValIleAlaProTrpArgMetProGIlyPhe 180
|||||
451 TACAACCGGTTCAAGGCGCGCATGACCTGATGAGTGAAGCAAGCAACA 500
|||||
181 TyrAsnArgPheLysGIlyArgAsnAspLeuMetGIlyAlaLysGIlnH 197
|||||
501 CGGATTCATCCATCCCGGTCACTCCCAAGAACCCGTGAGCATGATGAGA 550
|||||
197 sGIlyIleProIleProValThrProLysAsnProTrpSerMetAspGIuA 214
|||||
551 ACCTTCATGCATACAGTACGACAGAGCTGGAATCTCGAGAACCCCAAGAAC 600
|||||
214 snLeuMetHisIleSerTyrGIlyAlaGIlyIleuGIlyLysAsnProLysAsn 230
|||||
601 CAAGGCGCTCCAGGCTCTTACACGAGACCCAGAGACCCAGCAAGGCCCC 650
|||||
231 GlnAlaProProGIlyLeuTyrlThrlLysThrGlnAspProAlaLysAlaPr 247
|||||
651 CAACACCCCTGACATTCGAGATCGAGATCGAATCAAAAAGGGTCCCTGTGA 700
|||||
247 cAsnThrProAspIleLeuGIlyIleGIlyPheLysGIlyValProValL 264
|||||
701 AGGTGACCAACGTCAAGATGAGCAGCACCAGCAGCATCTCTTGAGACTC 750
|||||
264 ysvaIThrAsnValIysAspGIlyThrlHisGlnThrSerLeuGIleu 280
|||||
751 TTATGTTACTGTAACAGATGCGGGGAGCAGCATGCGTGGCCGTATGA 800
|||||
281 PheMeTyrlLeuAsnGIlyValAlaGIlyLysHisGIlyValGIlyArgIleAs 297
|||||
801 CATGTGAGAGACCGCTTCATTTGAATGAATGCCGAGGTATAGAGAG 850
|||||

```

```

297 pIleValGIuAsnArpHeIlleGlyMetLysSerArgGlyIleTyrGIuT 314
851 CCCGAGGAGGACCATCTTACCATGCTCATTTAACAATCGAGGCTTC 900
314 hPrIoAlaGIyThrIleuTyrHIsAlaHIsLeuSprIlleGIuAlaPhe 330
901 ACCATGAGCGGGAAGTGGCAAAATCAACAAAGCGCTGGCTTGAATT 950
331 ThMeTAsPArGIuValArGIuLysIlleGlnIleuGIuLysPhe 347
951 TGGTGAGCTGGTATACCGGTTTACGGCTTACGGCTGAGCTGATGAA 1000
347 eAlaGIuLeuValTyrThrGIyPheTrpHIsSerProGIuGlySerPhe 364
1001 TCCGCACTGCATCGCCAGTCCCAAGAGCGAGTGAAGGAAAGTGAC 1050
364 AlArGHIscYsIlleAlaLysSerGIuArGIuGIuLysValGIu 380
1051 GTTCCGCTCCCAAGGCGCAGGTATCATCTTCGGCGGAGTCCCACT 1100
381 ValSerValIleuLysGIuValArGIuLysGIuLysValSerProle 397
1101 GTCTCTACATAGAGAGCGGTGATGACATGAAGTGCAGGGTATATG 1150
397 uSerIleuTyrAsnGIuGIuLeuValSerMeTAsnValGIuGIuAspTyrG 414
1151 AGGCACTGATGCCACCGGTTTCATCAATCAATTCCTCAGCTGAG 1200
414 IuProThrsPArGIuThrGIyPheIlleAsnIlleAsnSerLeuArGIuLys 430
1201 GAATATCATCTCTCCAGAGCAAGTCACTGCCAA 1236
431 GIuTyrHIsArGIuGIuInseryValIthrAlaLys 442

seq.name: /cgn2_6/plodata/2/paa/PCTUS.COMB.pep:PCT-US00-26524B-6180

seq-documentation block:
: Sequence 6180, Application PC/TUS0026524B
: GENERAL INFORMATION:
: APPLICANT: Birste et. al.
: TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
: FILE REFERENCE: PA005PCT
: CURRENT APPLICATION NUMBER: PCT/US00/26524B
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: 60/163,280
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 6180
: LENGTH: 442
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (22)
: OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
PCT-US00-26524B-6180

alignment_scores:
  Quality: 2123.00      Length: 412
  Ratio: 5.178          Gaps: 0
  Percent Similarity: 99.515      Percent Identity: 99.272

alignment_block:
US-09-775-693-1 x PCT-US00-26524B-6180
Align seq 1/1 to: PCT-US00-26524B-6180 from: 1 to: 442
1 ATGTCCAGCAAGGCTCCGTGGTTCTGGCTACAGTGGCGGCTGGAGCAC 50
|||||

```

```

31 MetSerIeLysGIuSerValIValleuAlaTyrSerGIuGIuLeuAspTh 47
51 CTCTGCATCTCTGTGGTGAAGGACAGCATATGACATTCGCT 100
|||||
47 rSerCysIlleLeuValItrPheLysGIuGIuGIuGIuGIuGIuGIuGIu 64
101 ATCTGGCCCACTTGGCCAGAGAGACATTCGAGAGACCCGAGAGAG 150
|||||
64 yLeuAlaAsnIlleGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 80
151 GCACGTGAAGCTTGGGGCCCAAAAGGTTCATTGAGATTCACAGCAGA 200
|||||
81 AlAlaLeuLysLeuGIuAlaLysLysValPheIlleGIuAspValSerArg 97
201 GTTGTGGAGAGATTCATCTGGCCGCTTCATCCAGTCCAGGCGCATGAG 250
|||||
97 uPheValGIuGIuGIuPheIlleTrpProAlaIlleGIuInsSerAlaLeuTyr 114
251 AGGACCGCTACCTCTGGGACCTCTCTTGGCCAGGCGCTGCATCGCC 300
|||||
114 IuAspArgTyrIleuLeuGIuThrSerIleuAlaArgProCysIlleAlaArg 130
301 AAACAGTGGAAATCGCCACGCGGAGGGGCCAGTATGTGTCCACGG 350
|||||
131 LysGIuValGIuIleAlaGIuArgGIuGIuGIuGIuGIuGIuGIuGIu 147
351 CGCCACAGGAAAGGAGAGATCAGGTCGGCTTTCAGCTCAGCTGCTCT 400
|||||
147 yAlaThrGIuGIuLysGIuAsnAspGIuValArGIuPheGIuLeuSerCysTyr 164
401 CACTGGCCCGCCAGATAAGGTCAATTCCTCTCCGAGAGATGCTGAATTC 450
|||||
164 eIleuAlaProGIuIleLysValIleAlaProItrPArGIuMetProGIuPhe 180
451 TACACCGGTTCAAGCGCGCATGACCTGATGATGATGATGATGATGAT 500
|||||
181 TyrAsnArGIuPheLysGIuArgAsnAspLeuMetGIuTyrAlaLysGIu 197
501 CGGATTCCTCATCCCGGTCACTCCCAAGAACCCGCGGAGCATGATGAG 550
|||||
197 sGIuIleProIleProValIthrProLysAsnProItrPArGIuMetPro 214
551 ACCTCATGCACTACGATGAGAGGATGATGATGATGATGATGATGATG 600
|||||
214 snLeuMetHIsIleSerTyrGIuAlaGIuIleLeuGIuAsnProLysAsn 230
601 CAAGCGCTCCAGGCTCTTACACGAGAGACCCGAGGACCCCAAGCCCC 650
|||||
231 GIuAlaProProGIuLeuTyrThrLysThrGIuAspProAlaLysAlaPr 247
651 CAACACCCCTGACATTCGAGATCGAGTTCAAAAAGGGCTCCCTGTGA 700
|||||
247 oAsnItrProAspIleLeuGIuIleGIuPheLysGIuGIuGIuGIuGIu 264
701 AGGTGACCAAGCTCAAGATGGCACCCACCCAGACCTCTTGAAGTTC 750
|||||
264 yValAlaThrsAsnValLysAspGIuThrHIsGIuInsPheIleuGIuLeu 280
751 TTGATGTACCTGAGCAAGATCGCGGCAAGCATGGCTGGCGGTATTGA 800
|||||
281 PheMetCtyIleuAsnIleuValAlaGIuLysHIsGIuValGIuArgIleAs 297
801 CATGTGAGAACCGCTTCATTTGAGATGAAGTCCGAGCTATTCAGAGA 850
|||||
297 pIleValGIuAsnArpHeIlleGlyMetLysSerArgGIuIleTyrGIuT 314
851 CCCGAGGAGGACCATCTTACCATGCTCATTTAACAATCGAGGCTTC 900
|||||
314 hPrIoAlaGIyThrIleuTyrHIsAlaHIsLeuSprIlleGIuAlaPhe 330
901 ACCATGAGCGGGAAGTGGCAAAATCAACAAAGCGCTGGCTTGAATT 950
|||||
331 ThMeTAsPArGIuValArGIuLysIlleGlnIleuGIuLysPhe 347

```

```
seq_name: /cgn2_6/ptodata/2/paa/US099_COMB.pep:US-09-925-300-1594
seq_documentation_block:
```

```
seq_documentation_block:
  ! Sequence 1504
  ! ...
```

```

/ Sequence 1594, Application US/09925300
/ GENERAL INFORMATION:
/ APPLICANT: Craig Rosen,
/ APPLICANT: Steve Ruben
/ TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
/ FILE REFERENCE: PA101
/ CURRENT APPLICATION NUMBER: US/09/925,300
/ CURRENT FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: PCT/US00/05988
/ PRIOR FILING DATE: 2000-03-08
/ PRIOR APPLICATION NUMBER: 60/124,270
/ PRIOR FILING DATE: 1999-03-12
/ NUMBER OF SEQ ID NOS: 1890
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1594
/ LENGTH: 442
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (122)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ US-09-925-300-1594

```

	Quality:	Length:
Percent Similarity:	99.515	Gaps: 0
Percent Identity:	99.272	Coverage: 100%

Assignment_Block: US-09-775-693-1 x US-09-925-300-1594

Align seg 1/1 to: US-09-925-300-1594 from: 1 to: 442

1 ATATCGAAGAAAGGCTCCGTGGTTCTTGGGCTACATGGAGGCGCCGGACAC 50
31 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspHis 47
51 CTCTGTCAATCCTCGTGTGGCGTGAAGGACAAAGGCTATGAGCTATTGGCT 100
47 rSerGlyIleuValTyrPheuGlyIleuGlyTyrAspValIleuAlaT 64
101 AACTGGCCACATTGGCGAAGGAAGAACTCTGAGGAACCCGAGAAG 150

151	GCACGTGAAGCTTTGGGGCCCAAAAGGTGTTCATTGAGATGTGCACACGGA	200
81	AlaLeuLybLeuGlyAlaLysLysValPheLeuGluAspValSerArgGlu	97
201	GTTTGGGAGGAGTTTCATCTGGGCGGCATCGACGTCCAGCGACGTGATG	250
97	uPheValGluGluPheLeuPheLeuProAlaIleGlnSerSerAlaLeuArg	114
251	AGGACCGGTACCTCTGGGACACCTCTGTCCAGCGCCCTGCATCGCCGC	300
114	LuAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg	130
301	AAACAAGGGGAAATCCGCCACGCGGAGGAGGCGCCAAATATGTGCCACGG	350
131	LysGlnValGluIleAlaGlnArgGluGluAlaLysTyrValSerHisGlu	147
351	CGCCACAGGAAGGGGAAGCATGAGTCCGGTTTGAGTCCAGCTGTCTACT	400
147	YAlaThrGlyLysGlyLysAsnAspGluValAlaArgPheIleuSerCysArg	164
401	CACGTGACCCCGAGATAAAGGTCATTGGCTCCCTGGAGGATGCGTGAATTC	450
164	etiLeuAlaProGlnIleLysValIleAlaArgTrpArgMetArgGluPhe	180
451	TACAACCGGTGAAGGCCCGCAATGACCTGTAGTAGAGCAAAAGCAAC	500
181	TyrAsnAlaArgPheLysGlyLysAsnAspLeuMetGluTyrAlaLysGlnHis	197
501	CGGAGTTCCATCCCGGTCATCTCCCAAGAACCCGTGAGATGTGATGAGA	550
197	SeuIleProLeuProValThrProLysAsnProTrpSerMetAspGluAla	214
551	ACCTCATGACATAGGTAGAGGGTGGAAATCTGTGAGAACCCCAAGAAC	600
214	SnLeuMetHisIleSerTyrGlnAlaGlyIleLeuGluLysAsnProLysAsn	230
601	CAACGCGCTCCAGGTCTGTACACGAAGACCCAGACCCCAAGAACGCCCC	650
231	GlnAlaArgProGlyLeuLysTyrThrLysThrGlnAspProAlaLysAlaArg	247
651	CAACACCCCTGCACATCTCGAGATGAGCTGAAAAAGGGTCCCTGTGA	700
247	AsnThrProAspArgIleLeuGlnIleGluPheLysGlySeuIleProVal	264
701	AGGTGACAACTCAAGGATGGACACCCACACACCTCTTGGAGCTC	750
264	YValAlaThrAsnValLysAspArgLysThrThrHisGlnThrSerLeuGluLeu	280
751	TTCATGTACCTGCAGAACTGCGGCGACACATGAGGTGGCGCGATTTGA	800
281	PheMetTyrLeuAsnGluValAlaGluLysGlnValGluArgIleLys	297
801	CATGTGGAACAACGCTTCATTGGAATGAAGTCCGAGTATCTACGAGA	850
297	PuLeuValGluAsnArgPheLeuIleuMetLysSerAlaGlyIleTyrGluT	314
851	CCCGACGAGCGACCAATCCCTTACATGCTCATTTAGACATGAGGCGCTTC	900
314	HisProAlaGlyThrIleLeuLysThrHisAlaHisLeuAspArgIleGluAlaPhe	330
901	ACCATGACCGCGGAAGTGGCAAAATCAACAAGGCGCGGCTTGGAATTT	950
331	ThrMetAspArgGluValAlaArgLysIleLeuGlnIleuLysLeuSerHis	347
951	TGCTGAGCTGGGTATTCACCGGTTTAAAGCGCCATGACCTGAGTGAATTTG	1000
347	eAlaGluLeuValTyrThrGlyLysPheThrHisSerProGluCysGlyPheArg	364
1001	TCCGCGCATGTCATGCGCAAGTCCGAGGACGAGTGGAAAGGAGCGAG	1050
364	AlaArgHisCysThrAlaLysSerGluGlnLysArgValGluGlyLysValGln	380

```

1051 GTGTCGCTCAAGAGCCAGGTGATCCGCGCGGAGTCCCACT 1100
|||||
391 VALSERVALLEULYSGLYINLVALTYRILEULGLYARGGLUSERPROLE 397
|||||
1101 GTCTCTACATGAGAGCTGTGAGCATGAACGTGCAGGGTGATATG 1150
|||||
1151 AGCCAACTGATGCCACCGGTTTCATCAACATCATTCCTCGAGGTGAAG 1200
|||||
414 LUPROTTHRSPALATHRGYPHEILEASNILEASNISERLEUARGLEUYS 430
|||||
1201 GAATATCATGCTCCAGAGCAAGGCACTGCCAAA 1236
|||||
431 GLUTYRTHRSARGLEUINSERLYSVALTHRALLYS 442

seq_name: /cgn2_6/plodata/2/paa/US097_COMB.pep:US-09-760-475-2729

seq_documentation_block:
: Sequence 2729, Application US/09760475
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT49
: CURRENT APPLICATION NUMBER: US/09/760,475
: PRIORITY FILING DATE: 2001-01-16
: PRIOR APPLICATION DATA REMOVED - consult PAM or file wrapper
: NUMBER OF SEQ ID NOS: 4122
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2729
: LENGTH: 473
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (14)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (17)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-760-475-2729

alignment_scores:
Quality: 2123.00 Length: 412
Ratio: 5.178 Gaps: 0
Percent Similarity: 99.515 Percent Identity: 99.272

alignment_block:
US-09-775-693-1 x US-09-760-475-2729 ..

Align seg 1/1 to: US-09-760-475-2729 from: 1 to: 473

1 ATGTCCAGCAAGGCTCCGTGTTCTGGCCTACAGTGGCGGCTGGACAC 50
|||||
62 MetSerSerLysGlySerValValLeuAlaIyrSerGlyLeuAspTh 78
|||||
51 CTCTGCATCTCTGCTGGCTGAAGAAAGAGCTATGACGTATTCCT 100
|||||
78 rSerCysIleLeuValItrPleuLysGluGlnGlyTYRAspValIleIat 95
|||||
101 ATCTGGCCACATTTGGCCAGAGAGACTTTCGAGAGAGCCAGAGAGAG 150
|||||
95 yIleuAlaAsnIleGlyGlnLysGluAspPheGluAlaIaArgLysLys 111
|||||
151 GCACAGAGCTTTGGGGCCAAAAGTGTCATGAGATGCACAGAGGA 200
|||||
112 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerArg 128
|||||
201 GTTGTGAGAGATTCATCTGCGCGCCATCCAGTCCAGCGCAGCTGATG 250
|||||
128 uPheValIcIuGluIupheIleItrProAlaIleGlnSerSerAlaLeuTYR 145

```

```

251 AGGACCGCTACCTCTGGGCACTCTCTTCCAGAGCCCTGCATCCGCC 300
|||||
145 LuAspArgTYRLeuLeuGlyTHRSerLeuAlaArgProCysIleAlaArg 161
|||||
301 AAACAAGTGGAAATCGCCCAAGCGGAGGAGGAGCAAGATGATGCCAGG 350
|||||
162 LysGlnValGluIleAlaGlnIaArgGlnGlyAlaLysTYRValSerHisG 178
|||||
351 CGCCACAGGAAGAGGAGCATCAGTCCGCTTTCAGCTCAGCTGCTACT 400
|||||
178 yAlaThrGlyLysGlyAsnAspGlnValArgPheGlnLeuSerCysTYRS 195
|||||
195 eRleuAlaProGlnIleLysValIleAlaProItrPArgMetProGluPhe 211
|||||
451 TACAACCGGTTCAAGGGCGGCAATGACCTGTGAGATGACGAAGACACA 500
|||||
212 TYRAsnArgPheLysGlyArgAsnAspLeuMetCulTYRAlaLysGlnH 228
|||||
501 CGGATTCCTCCATCCCGGTCACTCCCAAGAACCCGTGAGCATGTGATAGA 550
|||||
228 SGLYLeuProIleProValThrProLysAsnProItrSerMetAspGluA 245
|||||
551 ACCTCATGCACATACGCTACAGAGCTGGAAATCCTGGAGAACCCCAAGAAC 600
|||||
245 snLeuMetHisIleSerTYRGlulAlaGlyIleLeuGluAsnProLysAsn 261
|||||
601 CAAGCGCTCCAGGCTCTCTACAGAGACCCAGAGACCAGCCAAAGCCCC 650
|||||
262 GlulAlaProProGlyLeuTYRThrLysThrGlnAspProAlaLysAlaPr 278
|||||
651 CAACACCCCTGACATTCGTGAGATCGAGTCAAAAAGGGTCCCTGTGA 700
|||||
278 oAsnThrProAspIleLeuGluIleGluPheLysGlyValAlaProValL 295
|||||
701 AGGTACCAACGCTCAAGATGGCACCACCCAGACCTCCTGTGAGAGTGC 750
|||||
295 ySValThrAsnValLysAspGlyItrThrHisGlnThrSerLeuGlnLeu 311
|||||
751 TTCATGTACCTGACAGAGTCCGGGCAAGCATGAGCGTGGCCGTATTGA 800
|||||
312 PheMetTYRLeuAsnGluValAlaIaGlyLysHisGlyValAlaArgIleAs 328
|||||
801 CATCGTGAAGAACCGCTTCATTTGGAATGAATGCCGAGATCTACAGAGA 850
|||||
328 rIleValGluAsnArgPheIleGlyMetLysSerArgGlyIleTYRGLUT 345
|||||
851 CCCAGCAGAGCACATCCCTTACCATGCTCATTTAGACATCGAGCCCTTC 900
|||||
345 hrProAlaGlyThrIleLeuTYRHisAlaHisLeuAspIleGluIaPhe 361
|||||
901 ACCATGGACCGGAAAGTGGCAAAATCAACAAGAGCTGGCTTGAATTT 950
|||||
362 ThrMetAspArgGluValArgLysIleLysGlnGlyLeuGlySerPhe 378
|||||
951 TGCTGAGCTGGTGTATACCGGTTTACGGCTAGCCCTGAGTGTGATTTT 1000
|||||
378 eAlaGluLeuValTYRThrGlyPheTrpHisSerProGluCysGluPheV 395
|||||
1001 TCCGACACTGCATCGCCCAAGTCCCAAGAGCAGTGAAGAGGAAGTGCAG 1050
|||||
395 aIaArgHisCysIleAlaLysSerGlnGluArgValGluGlyLysValGln 411
|||||
412 VALSerValLeuLysGlyGlnValTYRILEuGlyArgGluSerProle 428
|||||
1101 GTCTCTCAATGAGAGAGCTGTGAGCATGAGACGTGCAGGCGGATTTAG 1150
|||||
428 uSerLeuTYRAsnGluGluLeuValSerMetAsnValGlnGlyAspTYRG 445

```

1151 AGCCACTGATGCCCGGGTTCATCAACATCAATTCCTCAGCTGAAG 1200
 |||||||
 445 IuProthrasPalaThrglyPheIleasnIleasnSerIeuArGlyLys 461
 |||||||
 1201 GAATATCATGCTCCAGACGACGATCAGTCCGCAAA 1236
 |||||||
 462 GluTyHHisArGlyLeuGlnSerLysValThAlaLys 473
 |||||||

seq_name: /cgn2_6/ptodata/2/paa/US60_COMB.pep:US-60-207-216-791

seq_documentation_block:

; Sequence 791, Application US/60207216

; GENERAL INFORMATION:

; APPLICANT: Beasley, Ellen

; TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,

; FILE REFERENCE: CLO00580

; CURRENT APPLICATION NUMBER: US/60/207,216

; NUMBER OF SEQ ID NOS: 898

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 791

; LENGTH: 425

; TYPE: PRT

; ORGANISM: HUMAN

US-60-207-216-791

alignment_scores:

Quality: 1631.00

Ratio: 4.660

Percent Similarity: 84.951

Percent Identity: 80.825

Gaps: 2

Length: 412

US-09-775-693-1 x US-60-207-216-791

Align seg 1/1 to: US-60-207-216-791 from: 1 to: 425

1 ATGTCACGAAAGGCTCCGTTGCTGCTACAGTGGCGCTGAGAC 50
 |||||||
 50 MetSerSerLysGlySerValIleuAlaIlyrSerGlyLysLeuAspTh 66
 |||||||
 51 CCGGTCGACCTCCGTCGTCGAGGAAAGGCTATGAGTCAATGGCT 100
 |||||||
 66 rSerLysIleuLeuTrpLeuLysGlyLysIlyrValIleuAlaI 83
 |||||||
 101 ATCTGGCCACATGGCCAGAAAGACTTCAGAGAGCCAGAGAAAG 150
 |||||||
 83 yLLeuAlaAsnIleGlyLysIlyrValIleuAlaIlyrLys 99
 |||||||
 151 GCACGAAAGCTTTGGGCAAAAGGTTCATTTAGAGATGTCAGCAGGA 200
 |||||||
 100 AlaProLysLeuGlyThrLysLysValIleuAlaIlyrLys 116
 |||||||
 201 GTTGTGAGAGAGTTCATGTCGCGGCAATCCAGTCCAGCGCATGTATG 250
 |||||||
 116 uPheAlaGlnIleuLysIleuTrpAlaIleuLysSerSerLeuLys 133
 |||||||
 251 AGACACCGTACCTCTGGGCACTCTCTGCGAGGCGCTGATGCGCCGC 300
 |||||||
 133 IuAspGlyTyLLeuLeuGlyThrSerHisThrArGProCysThAlaArg 149
 |||||||
 301 AAACAAGTGGAAATCGCCAGCGGAGGGGCGCAAGTATGTGCCAGG 350
 |||||||
 150 Lys..... 150
 351 CGCCACAGAAAGGGGAGATCAGTCCGTTTACAGCTCAGCTGACT 400
 |||||||
 150 150
 401 CACTGCGCCCCGAGATTAAGGTCAATGCTCTCGAGAGATCGTGAATTC 450
 |||||||

151GlnIleuValIleAlaProTrpSerMetProGluPhe 163
 451 TAAACCGGTTCAAGGGCGCAATGACCTGATGAGTACGCAAGCAACA 500
 |||||||
 164 TyLAsnHisPheLysGlyCysSerArPheuThrgIuTyLThrgLysGlnI 180
 |||||||
 501 GGGGATTCGCAATCCCGGTCATCCCAAGAACCCGTGAGCATGGATGAGA 550
 |||||||
 180 sGlyLLeuProIleProValIleuTrpLysAsnProTrpAsnMetAspGlu 197
 |||||||
 551 ACCTCATGACATGACATGACAGGCTGGAATCTGGAGAA..CCCAAGAA 599
 |||||||
 197 snLeuMetHisIleSerTyGlyLysAlaGlyIleuGlnLysAspProGlnGlu 213
 |||||||
 600 CCAAGCGCTCCAGGCTCTTACAGAAAGCCAGAGCCAGCAAGCCAAAGCC 649
 |||||||
 214 ProSerAlaSerArGserLeuGlnIlyrProGly..LeuAlaLysAla 230
 |||||||
 650 CCAACACCCCTGACATCTCGAGATCGAGTTCAAAAAAGGGCTCCCTG 699
 |||||||
 230 ToAsnTrpProAspIleuGlnIlyrLeuGlnIlyrLysGlyValIleuLeu 246
 |||||||
 700 AAGGTGACCAAGTCAAGAGATGGCAGCAGCCAGCAGCAGCTCTGGAGCT 749
 |||||||
 247 LysValThrasValLysAspGlyThrThHisGlnThrSerLeuAlaLe 263
 |||||||
 750 CTTCATGTCCTGACAGAGTCCGCGGCAAGCATGGCGTGGCGCTAATG 799
 |||||||
 263 uPheMetTyLLeuAsnGlnValThrgIyLysHisCysValGlyArgIle 280
 |||||||
 800 ACATGTCGGAAGACCGCTCATGGAATGAAGTCCGAGGATCTACAGAG 849
 |||||||
 280 sPheValGlnLysAsnCysPheIleGlyMetLysSerArGlyTyLLeuGln 296
 |||||||
 850 ACCCCAGCAGCAGCAGCAGCTTACATGCTCATGTTGAGATCGAGCGCT 899
 |||||||
 297 ThrProAlaGlyThrgIleuTyLHisAlaHisIleuAspIleuIlyrTh 313
 |||||||
 900 CACCATGAGCCGGGAAGTGGGCAAAATCAAAACAGGCTGGGCTTAAT 949
 |||||||
 313 ethrMetAspTrpIleValArgLysIleuLysGlnIlyrLeuLysC 330
 |||||||
 950 TTGCTAGCTGCTGTATACCGGTTTACCGGCTTACCGGCTGAGTGAAT 999
 |||||||
 330 ysaLagIleuValTyLThrgIyPheTrpHisSerProLysGlnPhe 346
 |||||||
 1000 GTCCGCACTGTCATGCCCAAGTCCCAAGAGCAGTGAAGGAAAGTGA 1049
 |||||||
 347 ValArgHisSerIleAlaLysSerGlnIlyrGlyValGlyArgLysVal 363
 |||||||
 1050 GGTGTCCGTCCTCAAGGGCAGGTGACATGCTCCGCGCGGAGTCCCA 1099
 |||||||
 363 nValSerLeuPheLysAspGlnValTyLLeuGlyTyLTrpGlnSerPro 380
 |||||||
 1100 TGTCTCTTACAATGAGAGCTGTGAGCATGAGCTGAGGATGATAT 1149
 |||||||
 380 rGserLeuTyLAsnGlnIlyrLeuValSerMetAsnValHisGlyAspTr 396
 |||||||
 1150 GAGCCAGTACGATGCCAGCGGTTTCATCAACATCAATTCCTCAGCTGAA 1199
 |||||||
 397 GluProIleAspAlaThrgIyPheIleAsnIleAsnSerLeuArGlyLeu 413
 |||||||
 1200 GGAATATCATGCTTCCAGAGCAAGTCACTGCC 1233
 |||||||
 413 sGluTyHHisHisLeuGlnSerLysValThAla 424
 |||||||

seq_name: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:PCT-US01-08631-34811

seq_documentation_block:

; Sequence 34811, Application PC/TUS0108631

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

```

FILE REFERENCE: 21272-049
CURRENT APPLICATION NUMBER: PCT-US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO 34811
LENGTH: 433
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (112)..(157)
OTHER INFORMATION: Argininosuccinate synthase protein domain identified by
OTHER INFORMATION: EMATRIX, accession number BL00564C, p-value=1.000e-40, raw score
NAME/KEY: DOMAIN
LOCATION: (8)..(315)
OTHER INFORMATION: Argininosuccinate synthase domain identified by Pfam, accession
OTHER INFORMATION: name Argininosucc_synth, E-value=5.5e-151, Pfam score of 481.4
NAME/KEY: misc_feature
LOCATION: (1)..(433)
OTHER INFORMATION: xaa = X or * as defined in Table 2
PCT-US01-08631-34811

```

```

alignment_scores:
Quality: 1603.50      Length: 428
Ratio: 4.310          Gaps: 15
Percent Similarity: 86.916      Percent Identity: 80.140

```

```

alignment_block:
us-09-775-693-1 x PCT-US01-08631-34811 ..

```

```

Align seg 1/1 to: PCT-US01-08631-34811 from: 1 to: 433

```

```

1  ATGTCCAGCAAAAGGCTCCGTGCTTCTGAGCTACAGTGGCGCTGAGACAC  50
1  MetSerSerIysGlySerValValLeuAlaIyrSerGlyGlyLeuAspTh  17
51  CTCGTGATCTCTGCTGCTGAGCAAGCAAGGCTAGCTATGCTT  100
17  IserCysIleLeuValIrrPleuIysGlnGlnIlyrAspValIleAlaI  34
101  ATCTGGCCCAACTTGGCCAGAAAGAACTTCGAGAGGCCAGCAAGAG  150
34  yLeuAlaAsnIleGlyIlnIysGlnAspPheGlnIleAlaIlyrLys  50
151  GCACGTGAAGCTTGGGGCAAAAGGTTCATGTAGATGTACAGAGGA  200
51  AlaLeuIysLeuGlyAlaIlyslsValIrrPheIleGlnAspValSerArg  67
201  GTTGTGAGAGGATTCATCTGGCGCCATCCAGTCCAGCGCACTGTATG  250
67  upheValIglInupheIleIrrProAlaIleGlnIserSerAlaLeuIyrG  84
251  AGGACCGCTACTCTCTGGGACCTCTTGGCAGGCGCTGATGCCCGCC  300
84  IuAspArgIyrIleuLeuGlyIrrPhePheAlaIyrProcysIleAlaArg  100
301  AACCAAGTGAATCGCCAGCGAGGGGGCAAGTATGTGTCCACAGG  350
101  IysGlnValIglInIleAlaIlyrGlnIlyrAlaIlyrValSerHisGln  117
351  CGCCACAGAAAGGGAAAGATCAGTCCGCTGTTAGCTCAGCTGCTACT  400
117  yAlaThrGlyIysGlyAsnAspGlnValIArgPheGlnIleuSerCysIyr  134
401  CACTGGCCCCCAGATAAGGTGATTCCTCCGTGAGAGATGCTGAATTC  450

```

```

134  erLeuAlaProGlnIleIysValIleAlaIrrPrrArgMetProGlnPhe  150
451  TACAACGGCTTCAAGGGCCGCAATGACCTGTAGTACGCAAGCAAGA  500
151  TyrAsnArgPheIysGlyArgAsnAspLeuMetGlnIlyrAlaIysGln  167
501  CGGATTTCCATCCCGGTCTACTCCCAAGAACCCGTGAGCATGGATGAGA  550
167  sGlyIleProIleProValIrrProIysAsnProIrrSerMetAspGln  184
551  ACCCATGACATCAGCTACAGAGCTGAAATCCTGGAGAACCCCAAGAC  600
184  snLeuMetHisIleSerTyrGlnAlaIlyrIleuGlnsnProIlyAsn  200
601  CAAGCGCTCCAGGCTCTACAGAAAGACCCAGACCCCAAGACCC  650
201  GlnAlaIrrProGlyLeuIyrThrIlyrThsGlnAspProAlaIysAla  217
651  CAACACCCCTGACATTCGAGATCGAGTTCAAAAGGGGCTCCCTGTGA  700
217  oAsnThrProAspIleLeuGlnIleGlnPheIlyslsGlyValProVal  234
701  AG...GTGACCAACGTGACAGATGGCGACACCCAGACGACTCTGGAG  747
234  IuIlyGlyProThrPheIysAspGlyThrThrsGlnIrrPheLeuGln  250
748  CTCTCATGTACTGAAAGCAAGTGGCGGCAAGCATGGCTGGGC...CG  794
251  LeuPheMetIyrLeuAsnGlnIleValIlyrHisGlyAlaIlyrProIy  267
795  TATTGACATGTGGAGACCGCTTC...ATTGAAATGAATCCCGAGGA  841
267  rLeuThrSerTrpGlnAsnArgPheHisTrpGlnIleuIysSerArgGly  284
842  TCTAAGCAACCCCA...GCAGGACCATCCTTACCATGCTCATTTAGAC  888
284  IeLeuAlaIrrProGlnAlaGlyThrIleLeuIyrHisAlaHisIleuAs  300
889  ATCAGGCTTTCACCATGACCGGGAAGTGCGCAAAATC...AAACAGG  935
301  IleGlnAlaPheThrMetGlyIlyAspArgAlaGlnIleProAsnGln  317
936  CTGGGCTTGAATTTGCTGAGCTG...GTGTATACCGGTTTACGGCTA  982
317  yLeuGlyPheGlnPheValIleuGlnIlyrArgPheSerGlyThra  334
983  GCCCTGAGTGAATTTGCGGCACATCGCATCCGCAAGTCC...CAGGAG  1029
334  IArgGlnIysGlnIleuValIlyrProcysIleuArgGlnSerProGlnG  350
1030  CGAGTGAAGGGA...GTGCAAGTGTCCCTCTCAAGGCG...CAGGT  1073
351  ArgValIleuIlysserAlaIlyrAlProSerLeuIysGlyProArgCy  367
1074  GTACATCTCGCGCGGAGTCCCA...CTGTCTCTACATGAGAGAGC  1120
367  sThrSerLeuGlyProIlyrValProHisCysSerLeuIyrAsnGlnI  384
1121  TGTGAGACATGAACGTG...CAGGCTGATTAAGCAATCACTGAT...GCC  1164
384  euValIlyshIysGlyThrCysGlnIlyAspIyrGlnProAsn**LeuPro  400
1165  ACCGGTTCATCAACATCAATTCCTCAGCGTGAAG.....GATA  1205
401  ProGlyPheIleGln...ThrsIleSerLeuIysAlaIleuIlysnIy  416
1206  TCATGCTCTC...CAGAGCAAGGTCACTGCCA  1236
416  rHisArgLeuProArgAlaIlyrValThrAlaGln  427

```

```

seq_name: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:PCT-US01-08631-49199

```

```

seq_documentation_block:

```

```

; Sequence 49199, Application PC/TU050108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 49199
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (7)..(44)
; OTHER INFORMATION: Argininosuccinate synthase domain identified by
; OTHER INFORMATION: eMARTIX, accession number BL00564A, p-value=1.231e-27, raw score
; OTHER INFORMATION: 19.93
; NAME/KEY: DOMAIN
; LOCATION: (8)..(333)
; OTHER INFORMATION: Argininosuccinate synthase domain identified by Pfam, accession
; OTHER INFORMATION: name Arginosuc_synth, E-value=7.7e-145, Pfam score of 461.9
PCT-US01-08631-49199

```

alignment_scores:

```

Quality: 1431.50      Length: 407
Ratio: 4.618          Gaps: 2
Percent Similarity: 76.167      Percent Identity: 70.762

```

alignment_block:

```
US-09-775-693-1 x PCT-US01-08631-49199 ..
```

```
Align seg 1/1 to: PCT-US01-08631-49199 from: 1 to: 385
```

```

1 ATGTCCAGCAAGAGCTCCGTGCTTCTGCGCTACAGTGGCGGCTGACAC 50
|||||
1 MetSerSerLysGlySerMetValLeuAlaTyrSerGlyLeuAlaSnTh 17
51 CTGTCATCTCTGCTGCTGCTGAGCAAGACGATGACGTCATTGGCT 100
|||||
17 rSerCysIleLeuValTTrpLeuLysGlnGlnIlyTrpAspValThrAlaT 34
101 ATCTGGCCACATTTGGCCAGAGAGACTTGCAGAGACCAGAGAGAG 150
|||||
34 yLeuAlaSnThrGlyGlnGlyGlnGlyAspPheGlnGlnAlaArgLys 50
151 GCACCTAAGCTTGGGGCAAAAAGGTTCATTGAGAGATGTCAGCAGGA 200
|||||
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGlnAspValSerArgL 67
201 GTTTGGAGAGAGATTTCATCTGGCGGCGCATTCAGTCAGCGCATGTATG 250
|||||
67 urPheValGlnGlnPheIleTrpAlaAlaIleGlnSerSerAlaLeuTyrG 84
251 AGGACGCTACCTCTCTGGGACCTCTCTTGCAGGCGCTGCATCGCCGC 300
|||||
84 LysPheIleTrpLeuGlnGlyThrSerLeuThrArgProCysIleAlaHis 100
301 AAACAGTGAATCGCCCGAGGGAGGGGCGCAGTATGTGTCCACGAG 350
|||||
101 LysLysValGlnThrValGlnArgValGlyAlaLysTyrGlnGlyProT 117
351 CGGCACAGAGAAAGGAGACGATCAGGTCCGTTTGAGTCACAGCTGTACT 400
|||||
117 PArg..... 118
401 CACTGGCCCCCAGATAAAGTTCATTGCTCCTGAGAGATGCTGAATTC 450

```

```

119 ..... |||||
451 TACAACCGGTTCAGAGCGCCGACATGACCTGATGAGATACGAAGACACA 500
|||||
127 TyrAsnTrpPheLysGlyArgSerAspLeuMetLysTrpAlaLysGlnH 143
CGGATTCCTCCATCCCGGCTCACTCCCAAGAACCCCTGGAGCATGATGAGA 550
|||||
143 SGIyleProIleLeuValThrProLysAsnLeuTrpSerIleAspLysA 160
551 ACCTCAGTACATCAGCTACGAGGCTGGAATCTGGAAGAACCCCAAGAAC 600
|||||
160 snLeuMetHisIleSerTyrLysAlaGlyIleLeuGlnAsnProLysAsn 176
601 CAAGCGCTCCAGGTCTCTACAGCAAGACCCAGAGCCAGCAAGCCCC 650
|||||
177 GlnValProProGlyLeuTyrThrLysIleGlnAspProAlaLysAlaPr 193
651 CAACACCCCTGACATTCGAGATCGAGTTCAAAAAAGGGGCTCCCTGGA 700
|||||
193 AsnThrProAspIleLeuLysIleGlnPheLysLysGlyValProValA 210
701 AGGTGACCAACGTCAGAGATGGACCAACCAACGACCTCTTGAGACTC 750
|||||
210 rGValThrSerValLysAspGlyThrThrHisGlnThrSerLeuGlnLeu 226
751 TTCATGTACTGCAACGAAGTGGCGGCAAGATGCGCTGGGCGCTATTGA 800
|||||
227 PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValGlyTyrIleAl 243
801 CATCGTGAGAACCGCTTCATTGGAATGAAGTCCCGAGTATCTACGAGA 850
|||||
243 AlIleValGlnAsn..... 247
851 CCCCAGAGGACGACATCCCTTACCATGCTCATTTAGACATGAGGCTTTC 900
247 ..... 247
901 ACCATGGACCGGGAAGTGGCGCAAAATCAACACAGCCCTGGGCTTGAATT 950
247 ..... 247
951 TGCTGAGCTGTGTATACCGGTTACGCGCTAGCCCTGATGTGAATTG 1000
|||||
248 ..... IlePheThrArgGlnArgProSerProGlnCysLysLeuPheV 261
1001 TCCGCCACATGATGCGCAAGTCCAGAGAGCGAGTGAAGGAAAGTGCAG 1050
|||||
261 AlArgProCysIleAlaLysSerGlnGlnProAlaGlnGlyLysValGln 277
1051 GTGTCGCCGTCAGAGCGGCGGCTGATCATCTCGCGGAGAGTCCCATC 1100
|||||
278 ValProValLeuLysGlyGlnValTyrIleLeuGlyTrpGlnSerProLe 294
1101 GTCTCTACATAGAGAGCTGTGAGCATGAACAGTGCAGAGGTGATATTG 1150
|||||
294 uSerLeuTyrAsnGlnGlnLeuMetSerValAsnValGlnGlyAspTyrG 311
1151 AGCCAACTGATGCAACCGGCTTCATCAACATCAATTCCTCAGGCTGAG 1200
|||||
311 LysProIleAspAspThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 327
1201 GAATATCATGCTCTCCAGAGC 1221
328 GlnTyrHisCysLeuGlnSer 334
seq_name: /cgn2_6/plodata/2/paa/US60_COMB.pep:US-60-243-468-1235
seq_documentation_block:
; Sequence 1235, Application US/60243468
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen

```



```

: TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
: FILE OF INVENTION: AND USES THEREOF
: FILE REFERENCE: CL000929
: CURRENT APPLICATION NUMBER: US/60/243,468
: CURRENT FILING DATE: 2000-10-27
: NUMBER OF SEQ ID NOS: 2121
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 1235
: LENGTH: 317
: TYPE: PRT
: ORGANISM: HUMAN
: US-60-243-468-1235

```

```

alignment_scores:
  Quality: 1390.00      Length: 342
  Ratio: 4.680          Gaps: 4
  Percent Similarity: 86.842  Percent Identity: 81.287

```

alignment_block:

US-09-775-693-1 x US-60-243-468-1235 ..

Align seg 1/1 to: US-60-243-468-1235 from: 1 to: 317

```

1 ATGTCAGCAAGGCTCCGCTG...GTTCGGCTACAGTGGCGCTGGA 47
  |||||||
1 MetSerSerLysGlySerMetArgValLeuAlaTyrGlyGlyLeuAs 17
48 CACCTGTCATCTCTGCTGCTGAGAGACAGGCTATGACGCTATTG 97
  |||||||
17 pThrSerCysIleLeuValThrPalaLysGlnGlnGlyTyrAspValIleA 34
98 CCTATGTCGCAACATTTGGCCAGAGAGAGACTTGAGAACCCAGAG 147
  |||||||
34 LetyrLeuAlaAsnThrGlyGlnGlyGlnAspPheGlnGlnAlaArgLys 50
148 AAGGCCTAGAGCTGGGCAAAAGGTTCATTGAGATGTCAGACG 197
  |||||||
51 LysAlaLeuLysLeuGlyAlaLysLysValPheIleGlnAspValIleArg 67
198 GGAATTGTGAGAGAGTTCATCTGCGCGGCATCCAGTCCAGCGACTGT 247
  |||||||
67 gGluPheValGlnGluPheIleThrProAlaIleGlnSerThrLeuT 84
248 ATAGAGACCGCTACTCTCTGGGACCTCTTGGCCAGGCCCTGCATCGCC 297
  |||||||
84 TyrGlnAspArgTyrLeuLeuGlyThrPheLeuAlaArgProCysIleAla 100
298 CGCAACAGTGAATCGCCCGCGAGGAGGCGCAAGTATGTGTCCCA 347
  |||||||
101 SerLysGlnValGlnIleAlaGlnGlnGlnGly.....ValSerArg 114
114 gGlyAlaThrGlyGln.....TyrGlnValArgPheGlnLeuThrP 129
348 CGCGCGCACAGAAAGGGAACATCAGTCCGTTTGAAGCTCAGCTGCT 397
  |||||||
114 gGlyAlaThrGlyGln.....TyrGlnValArgPheGlnLeuThrP 129
398 ACTCAGTGGCGCCCGCAGATTAAGGTATGCTCTCTGGAGAGATCGCTGA 447
  |||||||
129 TyrSerLeuValProGlnIleLysGlyThrAlaProThrPargMetProGln 145
448 TTCTACAAAGGTTTCAAGGCGCCCAATGAGCTGATGAGATGCAAGA 497
  |||||||
146 PheTyrAsnArgPheLysGlyArgAsnAspLeuMetGlnTyrAlaLysGln 162
162 nHisGlyIleProIleProValThrProLysAsnProThrPargMetAsp 179
498 ACACGGGATTCCATCCGGTCACTCCCAAGAAACCCGTGAGCATGAGATG 547
  |||||||
548 AGAACCTCAGCAATCAGCTAGAGGCTGAGATCCTGAGAGAACCCCAAG 597
  |||||||
179 LysnLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLys 195

```

```

598 AACCAAGCGCTCCAGGCTCTCTACAGCAAGACCCAGACCCCAAGAC 647
  |||||||
196 AsnGlnValProProAsp.....
648 CCCCAACACCCCTGACATTTCTCGAGATCGAGTTCAAAAAGGGCTCCG 697
  |||||||
202 .....ArgValLysLysGlyValPro 209
698 TGAAGTGACCAAGCTCAGAGATGACACACACAGCAGACCTCTTGGAG 747
  |||||||
209 allLysValThrAsnValAlaAsnAspLysSerThrHisProThrSerLeuGln 225
748 CTCTTCATGTACCTGAAAGAGTCCGCGGCAAGCATGCGCGCTGAT 797
  |||||||
226 PhePheMetTyrLeuAsnGlnValAlaAspLysHisGlyValGlyCysThr 242
798 TGACATGTGTGAGAACCGCTTCATTTGATGAATGAGTCCCGAGTATCAG 847
  |||||||
242 rAspIleValGlnAsnArgPheThrGlyMetLysSerArgGlyIleTyrG 259
848 AGACCCGACGACGACCATCTTACCATGCTCATTTAGACATCGAGGCC 897
  |||||||
259 LuthrProAlaGlyThrIleLeuTyrHisThrHisLeuAspIleLysAla 275
898 TTACACATGAGACCGGAGAGTGCAGCAAAATCAACAGAGCTGGCTTGA 947
  |||||||
276 PheThrMetAspTrpGlnValCysLysIleLysGlnGlyLeuGlyLeu 292
948 ATTTCGTCAGCTGCTGATACCGGTTTACGCGCTAGCGCTGAGAGTGAAT 997
  |||||||
292 sPheAlaIleLeuValTyrThrAlaPheThrPHisSerProGlnCysGlu 309
998 TTGTCCGCACTGCATCGCCCAAGTCC 1023
  |||||||
309 heValArgHisCysIleAlaLysSer 317

```

seq_name: /cgn2.6/plodata/2/paa/US60_COMB_dep:US-60-212-664-486

seq_documentation_block:

```

: Sequence 486, Application US/60212664
: GENERAL INFORMATION:
: APPLICANT: Ladunga, Steve
: APPLICANT: Spier, Gene
: APPLICANT: Greenberg, Simon
: APPLICANT: Rabkin, Steven
: APPLICANT: Wang, Yu
: TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
: FILE REFERENCE: CL000687
: CURRENT APPLICATION NUMBER: US/60/212,664
: CURRENT FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 636
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 486
: LENGTH: 347
: TYPE: PRT
: ORGANISM: HUMAN
: US-60-212-664-486

```

```

alignment_scores:
  Quality: 1350.50      Length: 347
  Ratio: 4.502          Gaps: 3
  Percent Similarity: 86.455  Percent Identity: 78.386

```

alignment_block:

US-09-775-693-1 x US-60-212-664-486 ..

Align seg 1/1 to: US-60-212-664-486 from: 1 to: 347

```

1 ATGTCAGCAAGGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 50
  |||||||

```

```

1 MetSerSerLysGlySerIleValIleuAlaCysSerGlySerLeuAspThr 17
51 CTGTGATCCTGCTGTGGCTGAGAACAGCCTATGACGTATTGCT 100
17 rPheCysIleLeuValIleuLysGlySerLeuValIleuAla 34
101 ATCTGGCAACATTGGCCAGAACAGACTTCGAGAGAGAGAGAG 150
34 yIleuAlaSerIleGlyGlnLysGluAspPheGlnIleuAla 50
151 GCACCTGAGCTTGGGCGCAAAAAGGTTCATTGAGATGTCAGC 200
51 AlaLeuLysProGlyAlaLysLysMetLeuIleGluAspVal 67
201 GTTGTGAGAGAGTTCATCTGGCGGCGCATCCAGCTCAGCAGT 250
67 upheValIleGluIleuPheIleTyrProAlaMetGlnSer 84
251 AGGACCGCTACCTCTCTGGGACCTCTCTTCCAGGCTGACAGC 300
84 IuAspArgTyrIleuLeuGlyThrSerLeuAlaArgProCys 100
301 AAACAGTGAATAATGCCAGCGGAGGAGGCGGAGATGTGTCAC 350
101 LysGlnValIleGluValAlaGlnArgGlyAlaLysTyrVal 117
351 CGCCACAGGAAGAGGAGACGATCAGCTCGGTTTGAGTCAGC 400
117 yThrThrGlyLysGlyAsnAspGln..... 125
401 CACTGGCCCCCAAGATAAGTTCATGCTCCTCGAGAGATGCC 450
125 ..... 125
451 TACAACCGGTTCAAGGCGCCGCAATGACCTGATGAGTAGCA 500
126 .....GlyHisSerAspLeuThrGlyTyrAlaLysGln 137
501 CGGAGATTCCTCCAGCGGCTACTCCCAAGAACCTGAGCATG 550
137 gGlyIleProThrProValThrProLysAsnProTyrSerMet 154
551 ACCCTGACATGACATGACGAGGCTGAGATCTGAGAACCCCA 600
154 sLeuMetHisIleSerCysGlnAlaGlyIleLeuGlnAsn 170
601 CAAGGCGCTCCAGTCTCTACAGAGAACCCAGACCCAGCAAG 650
171 GlnAlaProProAlaLeuTyrThrLysThrGlnAspProAl 187
651 CAACACCCCTGACATTCGAGATCGAGTCAAAAAGGGTCCCT 700
187 GAsnThrProAspIleLeuGlnIleGluPheLysGlyVal 204
701 AGGTACCAACGTCAGAGATGGACCAACCCAGACCTCTCTGAG 750
204 yValThrAsnValLysAspSerAlaThrIleGlnThrSerLe 220
751 TTCATGATCTGAAACAGATCGCGGCAAGCATGGCGGCGGAT 800
221 PheMetTyrLeuAsnGluValAlaGlyLysHisSerValI 237
801 CATCTGGAGAACCGCTTCATTCGAGATGAGTCCGAGATTC 850
237 AlIleValIleGlnAsnArgPheIleGlyThrLysSerArg 254
851 CCGCAGCGAGCACCATCTTACAGATGCTTATGACATGAGAG 900
254 IeProAlaGlyThrIleLeuCysTyrAlaHisLeuAspIle 270
901 ACCATGACCGGAGAGTCCGCAAAATCAACAGGCTGAGCTT 950
271 ThrMetTyrArgGluValIleHisLysIleLysGlnLys 287

```

```

951 TGTGAGCTGTGTATACGGGTTTAC.....GGCCTAGCCCTGAGT 991
287 eAlaIleuValIleuValIleuValIleuValIleuValIleu 303
992 GTGATTTGTCCCGCAGTCATCGCAAGTCCCAAGAGC 1030
304 ValAsnLeuSerAlaThrAlaSerProSerProArgSer 316

```

seq_name: /cgn2_6/prodata/2/paa/US60_COMB pep:US-60-216-770-97

seq_documentation_block:

```

: Sequence 97 Application US/60216770
: GENERAL INFORMATION:
: APPLICANT: Ladunga, Steven
: APPLICANT: Spier, Eugene
: APPLICANT: Greenberg, Simon
: APPLICANT: Rabkin, Steven
: APPLICANT: Wang, Yu
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
: FILE REFERENCE: C1000727
: CURRENT APPLICATION NUMBER: US/60/216,770
: NUMBER OF SEQ ID NOS: 156
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 97
: LENGTH: 347
: TYPE: PRT
: ORGANISM: HUMAN
US-60-216-770-97

```

alignment_scores: Quality: 1350.50 Length: 347
Ratio: 4.502 Gaps: 3
Percent Similarity: 86.455 Percent Identity: 78.386

alignment_block: US-09-775-693-1 x US-60-216-770-97 ..

Align seg 1/1 to: US-60-216-770-97 from: 1 to: 347

```

1 ATGTCCAGCAAGGCTCCGTGCTTGGCTTACAGTGGCGCTGAGAC 50
17 rPheCysIleLeuValIleuLysGlySerLeuValIleuAla 34
51 CTGTGATCCTGCTGTGGCTGAGAACAGCCTATGACGTATTGCT 100
17 rPheCysIleLeuValIleuLysGlySerLeuValIleuAla 34
101 ATCTGGCAACATTGGCCAGAACAGACTTCGAGAGAGAGAGAG 150
34 yIleuAlaSerIleGlyGlnLysGluAspPheGlnIleuAla 50
151 GCACCTGAGCTTGGGCGCAAAAAGGTTCATTGAGATGTCAGC 200
51 AlaLeuLysProGlyAlaLysLysMetLeuIleGluAspVal 67
201 GTTGTGAGAGAGTTCATCTGGCGGCGCATCCAGCTCAGCAGT 250
67 upheValIleGluIleuPheIleTyrProAlaMetGlnSer 84
251 AGGACCGCTACCTCTCTGGGACCTCTCTTCCAGGCTGACAGC 300
84 IuAspArgTyrIleuLeuGlyThrSerLeuAlaArgProCys 100
301 AAACAGTGAATAATGCCAGCGGAGGAGGCGGAGATGTGTCAC 350
101 LysGlnValIleGluValAlaGlnArgGlyAlaLysTyrVal 117

```

```

351 CGCCACAGAAAGGGAACGATCAGTCCGGTTGAGCTAGCTGCTACT 400
    ::::::::::::::::::::::::::::
117 yThrThrclLysglYasnAspGln..... 125
401 CAGTGGCCCCCAGATTAAGTTCATGCTCCCTGGAGATGCTGAAATTC 450
125 ..... 125
451 TACAAACGGTTCAAGGCCGCAATGACCTGATGAGTACGCAAGAACACA 500
    ::::::::::::::::::::
126 ..... GlyHisSerAspLeuThrGlnAlaLysGlnAla 137
501 CGGATTTCCCATCCGGTCACTCCACAGAACCCGTGGACATGATGAGA 550
    ::::::::::::::::::::
137 gGlyLeuProThrProValThrProLysAsnProTyrSerMetAspGlu 154
551 ACCTCATGCACATTCAGCTACGAGGCTGGATCTGAGAACCCCAAGAAC 600
    ::::::::::::::::::::
154 snLeuMetHisIleSerCysGluAlaGlyIleLeuGluAsnProGluAsn 170
601 CAGGCCCTCAGGCTCTACACAGAACCCAGAACCCAGAACCCGCTGGA 700
    ::::::::::::::::::::
171 GlnAlaProProAlaLeuThrThrLysThrGlnAspProAlaLysThrPr 187
651 CAACACCCCTGACATTCCTGAGATGAGTTCAAAAAGGGGCTCCCTGGA 700
    ::::::::::::::::::::
187 oAsnThrProAspIleLeuGlnIleGlnPheLysGlyValProVal 204
701 AGGTGACCAACGTCAAGATGAGCACCACACCACTCTTGGAGCTC 750
    ::::::::::::::::::::
204 yValThrAsnValLysAspSerAlaThrHisGlnThrSerLeuGluLeu 220
751 TTCATGACCTGACAGAGTCCGGGCGACGATGGCGTGGCGGTTCGA 800
    ::::::::::::::::::::
221 PheMetThrLeuAsnGlnValAlaGlyLysHisSerValGlyAsnIleAl 237
801 CATGTGGAGAACCGCTTCATTTGGAATGAGTCCGAGGATCTAGAGAA 850
    ::::::::::::::::::::
237 aIleValGluAsnArgPheIleGlyThrLysSerArgGlyIleTyrArg 254
851 CCCAGAGGACCACTCTTACCATGCTATTAGATCGAGGCGCTTC 900
    ::::::::::::::::::::
254 laproAlaGlyThrIleLeuGlyTyrAlaHisLeuAspIleGluAlaPhe 270
901 ACCATGACCGGGAAGTCCGCAAAATCAACAAAGGCTGGGCTTGAAT 950
    ::::::::::::::::::::
271 ThrMetTyrArgGluValHisLysIleLysGlnCysArgGlyLeuLysPh 287
951 TGCTGAGCTGGTATACCGGTTTAC.....GGCCTAGCCCTGAGT 991
    ::::::::::::::::::::
287 eAlaGluLeuValLysThrGlyLysPheThrLeuSerGlyThrAlaLeuSer 303
992 GTGAATTTGCCCGCATGATCGCAAGTCCCAAGGAGC 1030
    ::::::::::::::::::::
304 ValAsnLeuSerAlaThrAlaSerProSerProArgSer 316

```

```

seq_name: /cgn2_6/pctdata/2/paa/pctus_COMB.pep.pct-US01-08631-49202
seq_documentation_block:
; Sequence 49202, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 49202

```

```

; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (23)..(60)
; OTHER INFORMATION: Argininosuccinate synthase protein domain identified by
; OTHER INFORMATION: EMATRIX, accession number BL00564A, p-value=2.800e-37, raw sco
; OTHER INFORMATION: 19.93
; NAME/KEY: DOMAIN
; LOCATION: (24)..(272)
; OTHER INFORMATION: Argininosuccinate synthase domain identified by Pfam, accession
; OTHER INFORMATION: name Argininosuc-synth, E-value=6.4e-68, Pfam score of 217.6
; NAME/KEY: misc_feature
; LOCATION: (1)..(487)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08631-49202

```

```

alignment_scores:
Quality: 1314.00 Length: 445
Ratio: 3.670 Gaps: 25
Percent Similarity: 80.449 Percent Identity: 69.438

```

alignment_block:

US-09-775-693-1 x PCT-US01-08631-49202 ..

Align seg 1/1 to: PCT-US01-08631-49202 from: 1 to: 487

```

1 ATGTCACGAAAGGCTCCGCGTTCGCTACAGTGGCGGCTGAGAC 50
    ::::::::::::::::::::::::::::::::::::
17 LeuSerLysGlnValThrLeuLysGlnGlnGlyTyrAspValIleAlaT 33
51 CTCGTGATCCTCGTGTGCTGAGAGACAGAGCTTACGATTCCT 100
    ::::::::::::::::::::::::::::::::::::::
33 rSerCysIleLeuValThrLeuLysGlnGlnGlyTyrAspValIleAlaT 50
101 ATTCGGCCACATTTGGCGCAAGAGAGACTTGCAGAGCCAGAGAG 150
    ::::::::::::::::::::::::::::::::::::::
50 yLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGlnAlaArgLys 66
151 GCACGTGAAGCTTGGGGCCAAAGAGTTCATTGAGATTCAGCAGGA 200
    ::::::::::::::::::::::::::::::::::::::
67 AlaLeuLysLeuGlyAlaLysValPheIleGluAspValSerArg 83
201 GTTGTGAGAGATTTCATCTGGCGGCGCATCCAGTCCAGCGCA...CTG 247
    ::::::::::::::::::::::::::::::::::::::
83 uPheValGlnGluPheIleTyrProAlaIleGlnValGlnAlaHisLeu 100
248 ATGAGAGACCGGCTCCTGGGC...ACCTCTCTTGCAGGCGCTGCATC 294
    ::::::::::::::::::::::::::::::::::::::
100 yGlnAspArgTyrLeuProGlnHisLeuSerLeuProGlyProCysIle 116
295 GCCGCCAAACAAGT...GAATGCGCCGAGGAGGCGCAAGTAT... 339
    ::::::::::::::::::::::::::::::::::::::
117 AlaArgLysGlnValGlyAsnSerProSerLysArgGlyAlaLysTyrVa 133
340 .GTGTCCACGCGGCGCACAGAAAGGAGAGATGAGTCCGTTTAGC 388
    ::::::::::::::::::::::::::::::::::::::
133 lValProThrGlyAlaThrGlyLysGlnGlnGlnValAlaArgPheGln 150
389 TCAGTGTACTACTGCGCCCGCCAGATTAAGTTCATGCTCCCTGAGG 438
    ::::::::::::::::::::::::::::::::::::::
150 euserCysTyrSerLeuAlaProGlnIleLysValIleAlaProTyrParg 166
439 ATGCCGATTTCTACACCGGTTCAAGGCGCATGATGAGTGA 488
    ::::::::::::::::::::::::::::::::::::::
167 MetProGluPheTyrAsnArgPheLysGlnGlnAsnAspLeuMetGluT 183
489 CGCAAGCAACAC...GGGATTCACCTCCGCTGCTCCCAAGAACCCGT 535
    ::::::::::::::::::::::::::::::::::::::
183 rGlyLysAlaThrLeuGlyPheProIleProValThrProLysAsnProT 200

```

```

536 GAGACATGATGAGAACCTCATGACATGACATGAGAGCTGGAATCCG 585
200 rpsrMetAspGluAsnLeuMetHisIleSerTyrGlnAlaIleLeu 216
586 GAGAGCCCAAGAACCAAGCGCTCCAGCTCTTACAGCAAGACCCAGA 635
217 GlnAsnProLysAsnGlnAlaProProGlyLeuTyrThrIleGlnI 233
636 CCCAGCAAGAGCCCAACACCCCT...GACATTCG...GAGATCGAT 679
233 yProSerGlnLysProProThrProLeuAspIleLeuGlyAspSer 250
680 TCAAAAAGAGGCTCCCTGGAAG...GTACCAACGTCAGAGATGCAAC 726
250 heLysGlyGlyAlaProValGlnGlyProThrPheLysAspGlyThr 266
727 ACCCAACAGACCTCCTGAGCTTTCATGACATGAGCAAGTGGGG 776
267 ThrHisGlnThrPheLeuGlnLeuPheMetTyrLeuAsnGlnValAlaG 283
777 CAGAGTCGCGCTGGGCGCTAT...GACATCGTGAAGACCGCTTCATG 823
283 yLysHisGlyArgAlaValAlaPheAspIleValGlnAsnArgPheIleG 300
824 GAATGAGTCC...CGAGTATTCAC...GAGACCCAGCAGAGC...ACC 864
300 yAsn***SerProArgGlyIleLeuAlaArgAlaGlyThrProAlaGlnHis 316
865 ATCCTTACCATGCTCATTTAGACATGAGCGCTTAC...ATGAGCG 911
317 SerPheTyrHisAlaHisLeuAspIleArgIleuSerProThrAsp 333
912 G...GAAGTGGCAAAATCAACAAGCGCTGGCTG...AAATTGGT 955
333 gGlySerGlySerIleLysGlnGlyProGlyAlaTyrPheAlaG 350
956 AGCTGCTATACCGCTTTA...CGGCTAGCCCTGAGTGAATTTGTC 1002
350 IuLeuValTyrThrGlyPheProAlaGlnAlaProGlnLysGluPhe... 365
1003 CGCCACATGC...ATCGCCAAATCCCAAGCGAGTGGAGAG 1040
366 .....CysProProThrGlyIleAlaLysValProArgLysProSer 380
1041 GAAAGTCCAGAGTCCCTCCCTCAAGGC.....CAGGTGTACA 1078
380 yLysGlyLysValGlnValAlaLysSerIleLysGlyProGlnValTyr 397
1079 TCCTC...GGCCGGAAGTCCCA.....CTGTCTCTACAAATGAGAG 1119
397 IeLeuSerGlyArgGlnIleProHisLysLeuLeuThrMetArgIle 413
1120 CTGTGAGCATGAGTGC...CAGGTGATTT...GAGCCAACTGATGC 1163
414 IeValLysHisGlyThrCysLysGlyAspTyrTrpArgProIleAsp 430
1164 CACCGGTTTCATCAATCATATTCCTCAGCTGAG...GAAT 1204
430 yHisArgLysSerIleHisGlnIleProPheLysAlaIleLysnT 447
1205 ATCATCGTCTC...CAGACCAAGTCACTGCCAA 1236
447 yHisArgLeuProArgAlaLysValThrCysGln 458

seq_name: /cgn2_6/plodata/2/paa/US60_COMB.pep:US-60-185-362-544
seq_documentation_block:
: Sequence 544, Application US/60185362
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Viven
: TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS.
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS.
: TITLE OF INVENTION: AND USES THEREOF

```

```

: FILE REFERENCE: CL000285
: CURRENT APPLICATION NUMBER: US/60/185,362
: CURRENT FILING DATE: 2000-02-28
: NUMBER OF SEQ ID NOS: 836
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 544
: LENGTH: 299
: TYPE: PRT
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)...(299)
: OTHER INFORMATION: Xaa = Any Amino Acid
US-60-185-362-544

alignment_scores:
Quality: 1207.50 Length: 300
Percent Similarity: 91.33 Gaps: 1
Percent Identity: 77.667

alignment_block:
US-09-775-693-1 x US-60-185-362-544
Align seg 1/1 to: US-60-185-362-544 from: 1 to: 299

316 GCCCAGCGGAGGGGGCAAGTATGTGTCACAGCGGCCACAGAGAGG 365
1 AlacIIntPrGlnGlyAlaLysTyrValSerHisSerIleMetClyLysol 17
366 GAGACATGAGTCCGCTTTCAGCTCAGCTGCTACTGAGCGCCCGACA 415
17 yAsnAspGlnValTyrPheGlyLeuAlaCysTyrSerLeuAlaProGlnI 34
416 TAAAGTCAATTCCTCCCTGAGGATGCTGAATTCATCAACCGGTTCAG 465
34 IeLysValIleAlaProGlyArgIleProGlnPheTyrAsnGlnSerLys 50
466 GGGCGCAATGACCTGATGAGTACGCAAGCAACAGCGGATCCCATCC 515
51 GlyArgSerAspLeuMetGlnTyrAlaGlnLysHisGlyIleProIlePr 67
516 GGTCTACCCCAAGAACCCGTGGAGCATGATGAGAACCTCATGACATCA 565
67 oValThrIleLysHisProThrAspMetAspGlnAsnLeuMetHisIle 84
566 GCTACAGAGCTGGATCTGTGAGAACCCCAAGAACAGCGCTTCAGGT 615
84 eHisGlnAlaGlyIleLeuGlnAsnProLysAsnGlnAlaProSerGly 100
616 CTCTACACGAGAACCCAGGACCCCAAGCGCCCAACACCCCTGACAT 665
101 IeTyrMetLysIleGlnAspLeuAlaLysAlaProAsnThrProAsnI 117
666 TCTCGAGATCGAGTTCAAAAAAGGGGCTCCTGTAAGGTGCAAGTCA 715
117 ePheLysThrGln...LysLysGlyValProValLysValThrSerIle 133
716 AGGATGAGCAACCCAGCAGACCTCTCTGAGCTTTCATGACTGTAAC 765
133 yAspGlyThrThrHisGlnThrSerLeuValLeuPheMetTyrLeuAsn 149
766 GAAGTGGCGGCAAGCATGCGCTGGGCGGTATTCACATCTGTGAGAACG 815
150 GluValAlaGlyLysHisSerValGlyHisIleAspIleValIleAsn 166
816 CTTCATTGGAATGAAGTCCGAGGTATCTACGAGACCCAGCAGCAGCA 865
166 gPheIleGlnMetLysSer**GlyIleCysLysThrProAlaGlyThrI 183
866 TCCCTTACCATGCTCATTTAGACATGAGCGCTTTCACATGAGCGGAA 915
183 IeLeuTyrHisProHisLeuAspIleGlnIlePheAlaMetGlnGlnI 199

```

```
916 GTGGCAAAATCAACAAGCCCTGGCTTGAAATTGCTGAGCTGTGTA 965
|||||
200 ValArgLysIleLysGlnGlyLeuGlyPheAlaGlnLeuValTy 216
966 TACGGTTTACGGCCTAGCCCTGAGTGTGATTTGTCGCCACATGCATCG 1015
|||||
216 rThrGlyPheTrpHisAsnProGlnCysAspPheAlaHisIscysIleA 233
1016 CCAAGTCCCAGAGCAGAGTGGAGGAAAGTGCAGGTGTCCGTCTCAAG 1065
|||||
233 lAlysSerGlnAspArgValGlnGlyValGlnValSerIlePheLys 249
1066 GGGCAGGTGTACATCTCTGGCCGGAGTCCCCACTGTCTCTACAATGA 1115
|||||
250 GlyGlnValTyrIleLeuGlyGlnGluProPheLeuSerLeuTyrSerG 266
1116 GGAGCTGTGTAGCATGAACGTGCAGGGGTGATTATGAGCCAACTGATGCCA 1165
|||||
266 yGlnGlnValSerMetCAsnValGlnGlyAsnAspGlnProValGlyAspT 283
1166 CCGGGTTCATCAACATCAATTCCTCAGGCTGAAGAAATATCATCGTCTC 1215
|||||
283 hrSerLeuIleAsnIleAsnSerLeuArGmetLysGluTyrHisHisLeu 299
```


OM of: US-09-775-693-1 to: Pending_Patents_AA_New:* out_format: pfs
Date: Feb 12, 2002 4:07 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame-n2p.model -DEV=rlp
-Q=/cgn2.1/USPTO.spool/US0975693/runat_12022002_124153_1396/app.query.fasta_1.1315
-DB=Pending_Patents_AA_New -QPM=fastan -SUFFIX=rapn
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000
-LOOEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELDP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blotsum62 -TRANS=human40.coi -LIST=45 -DOCLIGN=200
-THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US0975693 -CGN1_1_44 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-775-693-1
Query length: 1239
Database: Pending_Patents_AA_New:*
Database sequences: 216105
Search length: 44922408
Search time (sec): 443.280000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-981-353-87				2123.00	3133.67
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-340-187-701				2036.00	3000.91
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-634-187-1176				1603.50	2362.27
/cgn2.6/ptodata/2/paa/US10_NEM_COMB.pcp:US-10-015-127-9692				944.00	1384.36
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-79357				866.50	1268.85
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-79357				865.50	1268.34
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-55742				845.50	1237.67
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-55742				845.50	1237.16
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-897-516-5764				815.00	1199.49
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-11258				815.00	1190.76
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-675-784-12974				759.00	1114.64
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-79359				655.50	959.15
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-55744				639.50	935.38
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-11255				598.00	874.72
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-029-386-32991				477.00	702.75
/cgn2.6/ptodata/2/paa/US10_NEM_COMB.pcp:US-10-029-386-32991				477.00	702.18
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-78396				167.50	236.17
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-53190				160.50	225.79
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-5221				156.50	215.42
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-5220				156.50	213.65
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-5219				156.50	213.38
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-49072				153.00	212.97
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-55640				151.50	213.61
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-78930				147.50	207.51
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-78870				147.50	207.51
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-52153				147.00	205.34
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-53785				147.00	202.90
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-53785				147.00	202.55
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-53785				146.00	199.94
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-10-021-698-145				146.00	199.08
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-78956				145.00	203.34
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-78545				145.00	198.46
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-55291				144.50	203.05
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-55765				144.00	201.85
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-78361				144.00	196.29
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-53616				144.00	191.10
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-70031				140.00	197.25
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-70031				140.00	197.25
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-70029				140.00	196.98
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-55629				138.00	192.48
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-69207				136.50	190.97

/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-53125	136.50	183.93	0.0034
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-54261	135.50	184.27	0.0038
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-70389	134.00	184.27	0.0048
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-50632	133.00	185.41	0.0054
/cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-708-427-77679	132.50	187.74	0.0053

seq_name: /cgn2.6/ptodata/2/paa/US09_NEM_COMB.pcp:US-09-981-353-87

seq_documentation block:

Sequence 87, Application US/09981353
GENERAL INFORMATION:
APPLICANT: Lasek, Amy W.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981,353
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 412
TYPE: PRN
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No: 1543330CD1
US-09-981-353-87

alignment_scores:
Quality: 2123.00 Length: 412
Ratio: 5.178 Gaps: 0
Percent Similarity: 99.515 Percent Identity: 99.272

alignment block:

US-09-775-693-1 x US-09-981-353-87

Align seg 1/1 to: US-09-981-353-87 from: 1 to: 412

1	ATGTGTCACCAAGAGCTCCGTTGCTGCTACAGTGGCGCTGACAC	50
1	MetserSerlysglySerValValLeuAlarYrSerYglYleuAspTh	17
51	CTGTGTCATCTCTGTGTGCTGAGGACCAAGCTATACCTGCTGCT	100
17	rsercylsileuValtryleuYsglInglYtyrAspValleat	34
101	ATCTGTCACCAAGCTCCGTTGCTGCTGAGGACCAAGCTATACCTGCTGCT	150
34	YrLeuAlasnlleclYglInlYsglAspPhegluAlaYrlyls	50
151	GCACCTGAAGCTTGGGCGCAAAAGGTTCATTCAGATGTCACGAGGA	200
51	AlaleuYlsleuclYAlaleYlsValPhelelegluAspValserArgl	67
201	GTTTGTGAGGAGTTCATTCGCGCCGATTCAGTCCAGGCTGATG	250
67	uPheValgluInlPheletrProlAlleleInserleAlaleuYlg	84
251	AGGACCGTACCTCGGGGACCTCTGTCGACGCGCTGATCCGCC	300
84	lAspPrglyrleuLeuYglYrSerleuAlaYrPrcYsllelelary	100
301	AAACAGTGAATGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	350
101	YsglValGlnlAllelelAlaYrlylsValYrlylsValserHlgl	117
351	CGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	400
117	YrlylAlrlylsYsglYrlylsYrlylsYrlylsYrlylsYrlyls	134
401	CACTGGCCCCGAGTAAAGTTCATTCGCTGAGGAGTGCCTGAATTC	450

```

134 erleuAlaProGlnIleLysValIleAlaProTrpArgMetProGluPhe 150
451 TACAACGGGTCAAGGGCCCAATGACCTGATGGATGATACCAAGACACA 500
151 TyrSnrMrpPheLysGlyArgAsnAspMetGluTrpAlaLysGlnI 167
501 CGGAGTTCACATCCCGGTCACTCCCAAGAACCCGTGGACGATGATAGA 550
167 sGlyIleProIleProValThrProLysAsnProTrpSerMetCaspI 184
551 ACCATGACGACATGACGTACGAGGCTGGATCTCTGGAACCCCAAGAC 600
184 snLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLys 200
601 CAACGCGCTCCAGGTCTCTACAGCAAGACCCAGACCCCAAGCGCCC 650
201 GlnAlaProProGlyLeuTyrTrpLysThrGlnAspProAlaLysAl 217
651 CAACACCCCTGACATTCGAGATCGAGTTCAAAAAGGGTCCCTGTGA 700
217 oAsnThrProAspIleLeuGlnIleGluPheLysGlyValProValI 234
701 AGGAGACCAAGTCACAGGATGGACACACCCACAGACCTCTGGAGCTC 750
234 yValThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGlu 250
751 TTCATGTACCTGAACGAGTGGCGGCAAGCATGGCGTGGCGCTATGTA 800
251 PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValGlyArgIle 267
801 CATCGTGGAGAACCGCTCTCTTGGATGAAGTCCGAGGTACTACGAGA 850
267 pIleValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrG 284
851 CCCACAGACGACCATCTTACATGCTCATTTAGACATCGAGGCTTC 900
284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAla 300
901 ACCATGACCGGAGAGTGGCAAAATCAACAGGCTGGCTGTAATTT 950
301 ThrMetAspArgGlnValAlaArgLysIleLysGlnGlyLeuLys 317
951 TGCTGAGCGGTGTATACGGGTTTACGGCTTACCGCTGAGTGTGAAT 1000
317 eAlaGluLeuValTyrThrGlyPheTrpHisSerProGluLysGlu 334
1001 TCCGCCATGTCATCCGCCAAGTCCCAAGAGCGAGTGAAGGAAAGT 1050
334 aAlaArgHisCysIleAlaLysSerGlnGluArgValGlnGlyLys 350
1051 GTGTCCGCTCTCAAGGGCCAGGTGATACATCTTCGGCCGGGAGTCC 1100
351 ValSerValLeuLysGlnValTyrIleLeuGlyArgGlnSerProI 367
1101 GTCTCTCTCAATGAGAGCTGTGAGCATGAAGTGAAGGAGGTGATT 1150
367 uSerLeuTyrAsnGlnGluLeuValSerMetAsnValGlnGlyAsp 384
1151 ACCCAATGTATGCCACCGGGTTCATCAACATCAATTCCTCAGGTGA 1200
384 LuProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeu 400
1201 GAATATCATCGTTCACAGCAAGTCACTGCCAAA 1236
401 GlnTyrHisArgLeuGlnSerLysValThrAlaLys 412

```

seq_name: /cgn2_6/ptodata/2/paa/us60_NEW_COMB.pep:us-60-340-187-701

seq_documentation_block:

```

: Sequence 701, Application US/60340187
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Wang, Zhiwei

```

```

: APPLICANT: Wang, Jian-Rui
: APPLICANT: Ghosh, Malabika
: APPLICANT: Meng, Gezhi
: APPLICANT: Boyle, Bryan J
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: TITLE OF INVENTION: Polypeptides
: FILE REFERENCE: 813
: CURRENT APPLICATION NUMBER: US/60/340,187
: CURRENT FILING DATE: 2001-12-12
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03800
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 09/515,126
: PRIOR FILING DATE: 2000-02-28
: PRIOR APPLICATION NUMBER: US 09/577,409
: PRIOR FILING DATE: 2000-05-18
: PRIOR APPLICATION NUMBER: PCT/US01/04927
: PRIOR FILING DATE: 2001-02-26
: PRIOR APPLICATION NUMBER: US 09/519,705
: PRIOR FILING DATE: 2000-03-07
: PRIOR APPLICATION NUMBER: US 09/574,454
: PRIOR FILING DATE: 2000-05-19
: PRIOR APPLICATION NUMBER: PCT/US01/04941
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: US 09/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: US 09/649,167
: PRIOR FILING DATE: 2000-08-23
: PRIOR APPLICATION NUMBER: PCT/US01/08631
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: US 09/552,929
: PRIOR FILING DATE: 2000-04-18
: PRIOR APPLICATION NUMBER: US 09/770,160
: PRIOR FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: PCT/US01/08656
: PRIOR FILING DATE: 2001-04-18
: PRIOR APPLICATION NUMBER: US 09/577,408
: PRIOR FILING DATE: 2000-05-18
: PRIOR APPLICATION NUMBER: PCT/US01/14827
: PRIOR FILING DATE: 2001-05-16
: NUMBER OF SEQ ID NOS: 1192
: SOFTWARE: PL_FL_genes Version 6.0
: SEQ ID NO 701
: LENGTH: 583
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-340-187-701

```

alignment_scores:

```

Quality: 2036.00 Length: 397
Ratio: 5.168 Gaps: 0
Percent Similarity: 99.244 Percent Identity: 98.741

```

alignment_block:

US-09-775-693-1 x US-60-340-187-701 ..

Align seg 1/1 to: US-60-340-187-701 from: 1 to: 583


```

1  ATGTCCAGCAAGGCTCCGTGCTTCTGAGCTACAGTGGCGCTGAGAC 50
   |||||||
77  MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuSph 93
51  CTCGGCATCTCTCGTGGCTGAGCAAGCAAGGCTATGACGTCAATTCCT 100
   |||||||
93  rSerGlyLeuValTyrLeuLysGlyGlyGlyTyrAspValIleLeu 110
101  ATCTGCCACATTTGGCCGAGAGAAAGACTTCGAGAAAGCCAGAGAAAG 150
   |||||||
110  yLeuAlaAsnIleGlyGlnLysGlyAspPheGlyGlyAlaArgLys 126
151  GCAGTGAAGCTTGGGGCAAAAGGTCTTCATGAGATGTCACACAGGA 200
   |||||||
127  AlaleuLysLeuGlyAlaLysLysValPheIleGlyAspValSerArg 143
201  GTTTGGGAGAGATTCACTGAGCGCGCATCCAGTCCAGCGCACTGATG 250
   |||||||
143  upheValGlyGlyLeuPheIleTyrProAlaIleGlnSerAlaLeuTyr 160
251  AGGACCGCTACCTCCTGGGACCTCTCTTCAGGCGCTGCATGCGCCG 300
   |||||||
160  LuAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 176
301  AAGCAAGTGAATCCGCCAGGCGGAGGGGCCAAGTATGTGTCCACGG 350
   |||||||
177  LysGlnValGlnIleAlaGlnArgGlyGlyAlaLysTyrValSerHisG 193
351  CGGCACAGAAAGGAGACATCAGGTCGGCTTTGAGCTCAGCTCTACT 400
   |||||||
193  yAlaThrGlyLysGlyAsnAspGlnValArgPheGlyLeuSerCysTyr 210
401  CACTGGCCCCAGATTAAGTCAATGCTCCTCGAGAGATGCGCTGAATTC 450
   |||||||
210  eLeuAlaProGlnIleLysValIleAlaProTyrPheProGlyPhe 226
451  TACAACGGTTCAAGGGCCGCAATGACCTGATGAGATGACGAAGACACA 500
   |||||||
227  TyrAsnArgPheLysGlyArgAsnAspLeuMetGlyTyrAlaLysGln 243
501  CGGATTCCTCATCCGGTCACTCCCAAGAACCCGTGAGCATGATGAGA 550
   |||||||
243  sGlyIleProIleProValThrProLysAsnProTyrSerMetAspGln 260
551  ACCTCATGACATCAGTACGAGGCTGAGATCCTGGAGAACCCCAAGAAC 600
   |||||||
260  snLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLysAsn 276
601  CAAGCGGCTCCAGGTTCTACAGCAAGACCCAGACCCCAAGCCCC 650
   |||||||
277  GlnAlaProProGlyLeuTyrThrLysThrGlnAspProAlaLysAlaPr 293
651  CAACACCCCTGACATTCGAGATCGAGTTCAAAAAAGGGGTCCCTGTGA 700
   |||||||
293  CAsnThrProAspIleLeuGlnIleGlyPheLysGlyValAlaProVal 310
701  AGGTACCAAGTCAAGGATGGACCAACCCAGACCTCCTTGGAGCTC 750
   |||||||
310  yValThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGlyLeu 326
751  TTCATGTACTGAAGCAAGTGGCGGCAAGCATGGCTGGGCGCTGATTGA 800
   |||||||
327  PheMetTyrLeuAsnGlyValAlaGlyHisGlyValAlaArgIleAs 343
801  CATCGTGGAGAACCGCTTCATTTGATGAATGATCCGAGGTTCTACAGGA 850
   |||||||
343  rIleValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrGln 360
851  CCCGAGAGGACGACATCCTTACATGCTCATTTAGACATGAGAGCCCTTC 900
   |||||||
360  hProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGlyAlaPhe 376
901  ACCATGAGCCGGGAAGTGGCAAAATCAAAACAGGCTGGGCTTGAATTT 950

```

```

|||||
377  ThrMetAspArgGlyValAlaGlyLysIleLysGlnGlyLeuGlyLeuSph 393
951  TCGCAGGTGGTGTATACCGGTTTACGGCCTACGCCCTGAGTGTGAATTTG 1000
   |||||||
393  eAlaGlyLeuValTyrThrGlyPheThrPheHisSerProGlyCysGlyPhe 410
1001  TCCGCGCTGATCGCCCAAGTCCAGAGAGCGATGAGAGGAAAGTGCAG 1050
   |||||||
410  AlArgHisCysIleAlaLysSerGlnGlyArgValGlyGlyLysValGln 426
1051  GTGTCCGTCCTCAAGGCGCAGGTGATACCTCGGCGGAGTCCACAT 1100
   |||||||
427  ValSerValLeuLysGlyGlnValTyrIleLeuGlyArgGlySerProLe 443
1101  GTCCTCTCAATGAGAGAGCTGTGAGCATGACATGATGATGATGATG 1150
   |||||||
443  uSerLeuTyrAsnGlyGlyLeuValSerMetAsnValGlnGlyAspTyrG 460
1151  AGCCAAGTGAATCCACCGGTTTCATCAACATCATTCCTC 1191
   |||||||
460  LuProThrAspAlaThrGlyPheIleAsnIleLysPheLeu 473

```

seq_name: /cgn2_6/ptodata/2/paa/us60_NEW_COMB pep:us-60-340-187-1176

seq_documentation_block:

```

: Sequence 1176, Application US/60340187
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Wang, Zhiwei
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Ghosh, Malabika
: APPLICANT: Weng, Gezhi
: APPLICANT: Boyle, Bryan J
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: TITLE OF INVENTION: Polypeptides
: FILE REFERENCE: 813
: CURRENT APPLICATION NUMBER: US/60/340,187
: CURRENT FILING DATE: 2001-12-12
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-05-18
: PRIOR APPLICATION NUMBER: PCT/US01/04927
: PRIOR FILING DATE: 2001-02-26
: PRIOR APPLICATION NUMBER: US 09/519,705
: PRIOR FILING DATE: 2000-03-07
: PRIOR APPLICATION NUMBER: US 09/574,454
: PRIOR FILING DATE: 2000-05-19
: PRIOR APPLICATION NUMBER: PCT/US01/04941
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: US 09/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: US 09/649,167
: PRIOR FILING DATE: 2000-08-23
: PRIOR APPLICATION NUMBER: PCT/US01/08631

```

```

; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/770,160
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/US01/08656
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/14827
; NUMBER OF SEQ ID NOS: 1192
; SOFTWARE: PL_FIL_genes Version 6.0
; SEQ ID NO 1176
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(433)
; OTHER INFORMATION: Xaa = any amino acid or symbol as shown in the table 8 as set for
; US-60-340-187-1176

```

```

alignment_scores:
    Quality: 1603.50      Length: 428
    Ratio: 4.310          Gaps: 15
    Percent Similarity: 86.916    Percent Identity: 80.140

```

alignment_block:

US-09-775-693-1 x US-60-340-187-1176 ..

Align seg 1/1 to: US-60-340-187-1176 from: 1 to: 433

```

1 ATGTCACAGCAAGGCTCCGTTCTGCGCTACAGTGGCGGCTGAGAC 50
1 MetSerSerLysGlySerValIValLeuAlaIValSerGlyGlyLeuAspTh 17
51 CTCGGTCATCCTCGTGGTGGCTGAGAGCAAGGCTATGACGTACCTGCT 100
17 rSerGlyIleuValIValLeuLysGlyGlnGlyTyrAspValIleAlaI 34
101 ATTCGGCCACATTCGGCCAGAGAGACATTCGAGAGAGCCAGAGAGAG 150
34 YrLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGlnAlaIValLys 50
151 GCACGTGAAGCTTGGGGCCCAAAAGGTGTCATTGAGAGATGTCAGCAGA 200
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGlnAspValSerArgI 67
201 GTTGTGGAGAGATTTCATCTGGCGGCGCATCGATCCAGCGCATGATG 250
67 uPheValIGlnGlnPheIleTrpAlaIleGlnSerSerAlaLeuTyrG 84
251 AGGACCGCTACCTCTGGGACCTCTTTCGACGAGCGCTGATCGCCGCG 300
84 LuAspArgTyrLeuLeuGlyThrPhePheAlaArgProCysIleAlaArg 100
301 AACCACTGGAATCGCCGCGGAGGGGGGCGCAATATGATGTCACAGCG 350
101 LysGlnValIGlnIleAlaGlnArgGlnGlyAlaLysTyrValSerHisG 117
351 CGCCACAGAAAGGGAGACATCAGTCCGTTTGGCTGAGCTCAGCTCTCT 400
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGlnLeuSerCysTyrS 134
401 CACTGGCCCCCAGATAAAGGTCAATGCTCCCTGAGAGATGCTGAATTC 450
134 etLeuAlaProGlnIleLysValIleAlaProTrpArgMetProGlnPhe 150
451 TACACCGGTTCAAGGCGCAGATGACCTGATGAGTAGTACGCAAGCAACA 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGlnTyrAlaLysGlnH 167

```

```

501 CGGATTCACCGCGGTGCTACCTCCCAAGAACCCGTGAGCATGATGAGA 550
167 sGlyIleProIleProValThrProLysAsnProTrpSerMetAspGln 184
551 ACCTGATGACATCAGCTACGAGGCTGAGTGAATCTGAGAACCCCAAGAC 600
184 snLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLysAsn 200
601 CAAGCGCTCCAGGCTCTCTACACAGAACCCAGAGCCAGACCTCTTGAG 650
201 GlnAlaProProGlyLeuLysThrLysThrGlnAspProAlaLysAlaPr 217
651 CAACACCCCTGACATTCCTGAGATGAGTCAAAAAGGGTCCCTGGA 700
217 oAsnThrProAspIleLeuGlnIleGlnPheLysGlyValProValG 234
701 AG...GTGACCAAGCTCAAGATGCGACCCACCCAGACCTCTTGAG 747
234 IuGlyGlyProThrPheLysAspGlyThrThrHisGlnThrPheLeuGln 250
748 CTCTTCATGTCATCTGAGAGAGTGGCGGAGAGATGCGGTGGGC...CG 794
251 LeuPheMetTyrLeuAsnGlnValAlaIleLysHisGlyValGlyProTy 267
795 TATTGACATCGTGAGAACCGCTTC...ATTGAATGATGCCGAGGTA 841
267 rLeuThrSerTyrGlnAsnArgPheHisTrpGlnLeuLysSerArgGlyI 284
842 TCTACGAGACCCCA...GCAGGACACCTCTTACCATGCTCATTTAGAC 888
284 IeLeuArgArgProGlnAlaIleLysThrLeuLysThrHisAlaHisLeuAsp 300
889 ATCGAGGCTTCACCATGAGACCCGAGAGTGGCAAAATC...AAACAGG 935
301 lIleGlnAlaPheThrMetGlyGlyAspArgAlaGlnIleProAsnGlnI 317
936 CCTGGGCTTGAATTCGTGAGCTG...GTGTTACCGGTTTACGGGCTTA 982
317 YLeuGlyIlePheGlnPheValGlnLeuGlyValTyrArgPheSerGlyThr 334
983 GCCCTGAGTGAATTCCTCCGACATCGATCCCAAGTCC...CAGSAG 1029
334 lApGlnLysGlnLeuValGlyProCysLeuArgGlnSerProGlnGln 350
1030 CGAGTGAAGAGAA...GTGACAGTGTCCGCTCTCAAGGCG...CAGGT 1073
351 ArgValGlnLysSerAlaGlyValProSerLeuLysGlyProArgCys 367
1074 GTACATCTCGCGCGGAGTCCCA...CTGCTCTCTTCATGAGAGAC 1120
367 sThrSerLeuGlyProGlnValProHisCysSerLeuLysGlnGln 384
1121 TGGTGAGCATGAAGCTG...CAGGGTGAATTAAGCAACTGAT...GCC 1164
384 euValLysHisGlyThrCysGlnGlyAspTyrGlnProAsn**LeuPro 400
1165 ACCGGGTATCATCAATTCATCCCTCAGGCTGAG...GAATA 1205
401 ProGlyPheIleGln...ThrSerIleSerLeuLysAlaGlnGlyAsnTy 416
1206 TCATCGTCTC...CAGAGCAAGGTCACTGCCAA 1236
416 rHisArgLeuProArgAlaLysValThrAlaGln 427

```

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB pep-US-10-015-127-9692

seq_documentation_block:

```

; Sequence 9692, Application US/10015127
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

```


LOCATION: 1..425
OTHER INFORMATION: Ceres Seq. ID 1965189
US-09-708-427-79358

alignment_scores:
Quality: 866.50 Length: 393
Ratio: 2.947 Gaps: 5
Percent Similarity: 74.809 Percent Identity: 45.547

alignment_block:
US-09-775-693-1 x US-09-708-427-79358 ..

Align seg 1/1 to: US-09-708-427-79358 from: 1 to: 425

```

19 GGTGCTGCTGCTACAGTGGCGGCTTGAGACACCTGTCATCTGCTGTG 68
   |||||.....|.....|.....|.....|.....|.....|.....|
30 ValValLeuAlaIleTyrSerGlyLeuAspThrSerValIleValProtr 46
   |||||.....|.....|.....|.....|.....|.....|.....|
69 GGTGAAGAACAA...GGCTATGACGTATGCTTATGCGCAACATG 115
   |||||.....|.....|.....|.....|.....|.....|.....|
46 PleuAtgGluAsnTyrGlyCysGluValAlaCysPheThrAlaAspValG 63
   |||||.....|.....|.....|.....|.....|.....|.....|
116 GCCAGAACAA...GACTTCGAGAGAACGACGAGGAGTTCGAGAGA 162
   |||||.....|.....|.....|.....|.....|.....|.....|
63 LylGlnGlyAlaIleGluLeuGlnGlyLeuGlnLysAlaLysAlaSer 79
   |||||.....|.....|.....|.....|.....|.....|.....|
163 GGGGCCAAAGAGTTCATGAGAGATGTCAGCAGGAGTTCGAGAGA 212
   |||||.....|.....|.....|.....|.....|.....|.....|
80 GlyAlaCysGlnLeuValAlaLysAspLeuLysGlnGluPheValSerI 96
   |||||.....|.....|.....|.....|.....|.....|.....|
213 GTTCATCTGGCGGCGCATCCAGTCGACGACGCTGATGAGACCGCTACC 262
   |||||.....|.....|.....|.....|.....|.....|.....|
96 uTyrIleTyrProCysLeuArgAlaGlyAlaValTyrGlnAlaTyrLys 113
   |||||.....|.....|.....|.....|.....|.....|.....|
263 TCCGCGGACCTCTCTTCCAGGCGCTGATCCCGGCAACAGTGGAA 312
   |||||.....|.....|.....|.....|.....|.....|.....|
113 euLeuGlnTyrSerMetAlaArgProValIleAlaLysAlaMetValAsp 129
   |||||.....|.....|.....|.....|.....|.....|.....|
313 ATCCGCCAGGCGGAGGGCCAGTATGTGTCCACGCGGCCACAGAGAA 362
   |||||.....|.....|.....|.....|.....|.....|.....|
130 ValAlaLysGlnValAlaGlyAlaAspAlaValAlaHisGlyCysThrGly 146
   |||||.....|.....|.....|.....|.....|.....|.....|
363 GGGAGGATGACGCTCGGCTTGGAGCTGACGCTGCTACTGAGGCGCCGC 412
   |||||.....|.....|.....|.....|.....|.....|.....|
146 sGlyAsnAspGlnValAlaArgPheGluLeuThrPheTyrAlaLeuAsnPro 163
   |||||.....|.....|.....|.....|.....|.....|.....|
413 AGATTAAGGTCATGCTCCCTGAGAGATGCTGAATTCACACCGGCTTC 462
   |||||.....|.....|.....|.....|.....|.....|.....|
163 IuLeuLysValAlaIleProtrPArgGluTyrPasp.....Ile 175
   |||||.....|.....|.....|.....|.....|.....|.....|
463 AAGGCGCGAATGACCTGATGAGTACGAAAGCAACAGGAGTTCATC 512
   |||||.....|.....|.....|.....|.....|.....|.....|
176 ThrGlyArgGluAspAlaIleGluTyrAlaLysLysHisAsnValProva 192
   |||||.....|.....|.....|.....|.....|.....|.....|
513 CCCGGTCACTCCAGAACCCGTGAGCATGAGAGAACCTCATGACA 562
   |||||.....|.....|.....|.....|.....|.....|.....|
192 IProValSerLysLysSerIleTyrSerArgAspArgAsnLeuThrPstL 209
   |||||.....|.....|.....|.....|.....|.....|.....|
563 TCAGCTACGAGGCTGATCTCGGAGAACCCCAAGAACCAAGCCCTCCA 612
   |||||.....|.....|.....|.....|.....|.....|.....|
209 euSerHisGlnGlyLysPheLeuGlnAspProAlaAsnGluProLysGlu 225
   |||||.....|.....|.....|.....|.....|.....|.....|
613 GGTCTCTACAGAACCCAGACCCAGCCAAAGCCCGCCAGACCCCTGA 662
   |||||.....|.....|.....|.....|.....|.....|.....|
226 AspMetTyrMetSerIleAlaProGluAsnAlaProSerLysProGlu 242
   |||||.....|.....|.....|.....|.....|.....|.....|
663 CATTCCTCAGATCGATTCAAAAAGGGTCCCTGTGAAGGTGACCAAG 712
   |||||.....|.....|.....|.....|.....|.....|.....|
242 uTyrLeuGlnIleGlyIleIleAlaGlyValProValSerIle...Asn 258
   |||||.....|.....|.....|.....|.....|.....|.....|
713 TCAAGATGCGACACCCAGACCTCTGAGAGCTTCATGATGATCAG 762
   |||||.....|.....|.....|.....|.....|.....|.....|

```

```

258 LysArgAsp.....LeuSerProAlaSerLeuLeuAlaGluLeu 270
763 AAGGAAGTCGCGGCGCAAGCATGCGCTGGCGCTTATGACATGCTGAGAA 812
   |||||.....|.....|.....|.....|.....|.....|.....|
271 AsnGlnIleGlyGlyLysHisGlyIleGlyArgIleAspMetValGluAs 287
   |||||.....|.....|.....|.....|.....|.....|.....|
813 CCGCTCATTTGATGAGTCCCGAGGTATCTACAGACACCCAGCAGGCA 862
   |||||.....|.....|.....|.....|.....|.....|.....|
287 nArgLeuValCysMetLysSerArgGlyValTyrGlnThrProGlyLys 304
   |||||.....|.....|.....|.....|.....|.....|.....|
863 CCATCTTTACCATGCTCATTTAGACATGAGGCGCTTCACCATGAGCCG 912
   |||||.....|.....|.....|.....|.....|.....|.....|
304 hIleMetAlaIleAlaValAlaArgGlnLeuGlnSerLeuThrLeuAsp 320
   |||||.....|.....|.....|.....|.....|.....|.....|
913 GAAGTCGCGCAATCAACAAAGCGCTTGGAGTTCGATGCTGCTG 962
   |||||.....|.....|.....|.....|.....|.....|.....|
321 GlnThrMetGlnThrLysAspIleIleAlaLeuLysTyrAlaGlnLeuVa 337
   |||||.....|.....|.....|.....|.....|.....|.....|
963 GTATACCGGTTTACGCGCTTACCGCTGAGTGTGAATTTGTCCGCACTGA 1012
   |||||.....|.....|.....|.....|.....|.....|.....|
337 IlyrAlaGlyArgTyrPheAspProLeuArgGlnSerPheAspAlaPheM 354
   |||||.....|.....|.....|.....|.....|.....|.....|
1013 TCGCCAGTCCCGAGGAGGAGTGAAGAGAAAGTGCAGGTGCTGCTTC 1062
   |||||.....|.....|.....|.....|.....|.....|.....|
354 eGluLysValThrAlaThrThrThrGlySerValThrLeuLysLeuTyr 370
   |||||.....|.....|.....|.....|.....|.....|.....|
1063 AAGGCGAGGTGTATCATCTCGCGCGGAGTCCCGCTGCTGTCTACAA 1112
   |||||.....|.....|.....|.....|.....|.....|.....|
371 LysGlySerValAsnValAlaSerArgLysSerProTyrSerLeuTyr 387
   |||||.....|.....|.....|.....|.....|.....|.....|
1113 TGAGAGAGCTGTGAGCATGACGTCGAGGCTGATATGAGCCAGCTGATG 1162
   |||||.....|.....|.....|.....|.....|.....|.....|
387 gGluAspIleSerSerPheGlnAsnGlyIleTyrAsnGlnAlaAspA 404
   |||||.....|.....|.....|.....|.....|.....|.....|
1163 CCACCGGTTTCATCAATCATATTCCTC 1191
   |||||.....|.....|.....|.....|.....|.....|.....|
404 IagGlnGlyPheIleArgLeuTyrGlyLeu 413

```

seq_name: /cgn2.6/plodata/2/paa/us09_NEW_COMB.pep:US-09-708-427-79357

```

seq_documentation_block:
; Sequence 79357, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79357
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..446
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..446
; OTHER INFORMATION: Ceres Seq. ID 1965188
; US-09-708-427-79357

```

alignment_scores:

Quality: 866.50 Length: 393
Ratio: 2.947 Gaps: 5
Percent Similarity: 74.809 Percent Identity: 45.547

alignment_block:

US-09-775-693-1 x US-09-708-427-79357 ..

Align seg 1/1 to: US-09-708-427-79357 from: 1 to: 446

```

19 GTGGTTTGCCCTACAGTGGCGCCCTGGACACCTGTCGATCCCTGCTG 68
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 ValValIleuAlaIleuIleuIleuIleuIleuIleuIleuIleuIleu 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 GCTGAAGGACAA...GGCTATGACGTCAATGGCTATGCTGGCCCAACAT 115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 pleuArgIleuAsnIleuIleuIleuIleuIleuIleuIleuIleuIleu 84
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 GCCAGAGGAA...GACCTGCGAGAGGAGGAGGAGGAGGAGGAGGAGG 162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 LylGlnGlyAlaIleuIleuIleuIleuIleuIleuIleuIleuIleu 100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 GGGGCGCAAAAGGTTCATTTAGAGATGACAGGAGGAGGAGGAGGAGGAG 212
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 GylAlaCysGlnIleuValIleuIleuIleuIleuIleuIleuIleuIleu 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 GTTCATTCGGCGCCGATCCAGTCCAGGCGGAGTATGAGACCCCTACC 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 uTrIleuIleuProCysIleuAlaGlyAlaValIleuIleuIleuIleu 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 TCCTGGGCGCTCTCTGCGAGGCGCCGATGCGCCGCAACAGTGGAA 312
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 euIleuGlyThrIleuIleuIleuIleuIleuIleuIleuIleuIleu 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 ATGCGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 ValAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 167
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 GGGGAGACATCAGTCCGCTTGGAGTCACTGCTCACTGCTGCTGCTGCT 412
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 sGlyAsnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413 AGATAAAGGTCAATGCTCTGCGAGGAGGAGGAGGAGGAGGAGGAGG 462
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 IuIleuIleuValIleuIleuIleuIleuIleuIleuIleuIleuIleu 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
463 AAGGCGCGCATGACCTGATGAGTACGCAACAGGAGGAGGAGGAGGAG 512
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 ThrGlyArgGlnAlaIleuIleuIleuIleuIleuIleuIleuIleuIleu 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 CCGGCTCACTCCAGAACCCCTGAGATGATGATGATGATGATGATGAT 562
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 IProValSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
563 TCAGCTACGAGGCTGATCTGAGAGACCCCAAGACCAAGGAGGAGGAG 612
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
230 euSerHisGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 246
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 GGTCTCTACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 662
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
247 AspMetIleuMetSerIleuIleuIleuIleuIleuIleuIleuIleuIleu 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 CATTCCTGAGATCGATCGATCGATCGATCGATCGATCGATCGATCGAT 712
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 uTrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 279
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
713 TCAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 762
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 LysArgAsp...LeuSerProIleuIleuIleuIleuIleuIleuIleu 291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 AAGGAGTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 AsnGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
813 CCGGCTCACTGAGATCGATCGATCGATCGATCGATCGATCGATCGAT 862
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 uArgIleuValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 325
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
863 CCATTCCTTACCATGCTCAATTTAGACATGAGGAGGAGGAGGAGGAGGAG 912
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 hTrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 341
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

913 GAAGTGGCAAAATCAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 962
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 GluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 358
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
963 GTATACCGGTTTACGCGCTAGCCCTGAGTGTGAATTTGTCCGCGCATG 1012
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
358 LTrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 375
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1013 TCAGGAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1062
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 euGlyIleuValIleuIleuIleuIleuIleuIleuIleuIleuIleu 391
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1063 AAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392 LysGlyIleuValIleuIleuIleuIleuIleuIleuIleuIleuIleu 408
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1113 TGAGGAGTGTGAGTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 gGlyAspIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 425
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1163 CCACCGGTTTCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1191
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 IArgIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 434
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /cgn2_6/plodata/2/paa/US09_NEW_COMB.pep:US-09-708-427-55743

seq_documentation_block:

```

? Sequence 55743, Application US/09708427
? GENERAL INFORMATION:
? APPLICANT: N. ALEXANDROV et al.
? TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
? FILE REFERENCE: 2750-1243P
? CURRENT APPLICATION NUMBER: US/09/708,427
? NUMBER OF SEQ ID NOS: 85364
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 55743
? LENGTH: 426
? TYPE: PRT
? ORGANISM: Zea mays subsp. mays
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1..426
? OTHER INFORMATION: Xaa is any amino acid
? NAME/KEY: misc_feature
? LOCATION: 1..426
? OTHER INFORMATION: Ceres Seq. ID 1935962
US-09-708-427-55743

```

alignment_scores:

```

Quality: 845.50 Length: 389
Ratio: 2.936 Gaps: 6
Percent Similarity: 74.036 Percent Identity: 45.501

```

alignment_block:

US-09-775-693-1 x US-09-708-427-55743 ..

Align seg 1/1 to: US-09-708-427-55743 from: 1 to: 426

```

19 GTGGTTTGCCCTACAGTGGCGCCCTGGACACCTGTCGATCCCTGCTG 68
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 Val***LeuAlaIleuIleuIleuIleuIleuIleuIleuIleuIleu 46
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 GCTGAAGGACAA...GGCTATGACGTCAATGGCTATGCTGGCCCAACAT 115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46 pleuArgIleuAsnIleuIleuIleuIleuIleuIleuIleuIleuIleu 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 GCCAGAGGAA...GACCTGCGAGAGGAGGAGGAGGAGGAGGAGGAGG 162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 LylGlnGlyAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 79
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

163 GGGCCAAAAGGTGTCATTCAGATGTCAGGAGGAGTGTGGAGGA 212
||||| : : : : : : : : : : : : : : : : : : : : : : : :
80 GLYAlAcysGlnLeuValValLysAspLeuLysGluGluPheValSerGI 96
||||| : : : : : : : : : : : : : : : : : : : : : : : :
213 GTTCATCTGGCCGGCCATCCAGTCAGCCGACTGTATGAGGACCGCTACC 262
||||| : : : : : : : : : : : : : : : : : : : : : : : :
96 uTYrIleTYrProCysLeuArgIaGlyAlaValTYrGluArgLYSTyTL 113
||||| : : : : : : : : : : : : : : : : : : : : : : : :
263 TCCTGGGACCTCTCTGGCAGCCCTGCATGCGCCGCAACAGAGTGA 312
||||| : : : : : : : : : : : : : : : : : : : : : : : :
113 euleuGlyThrSerMetAlaArgProValIleAlaLysAlaMetValAsp 129
||||| : : : : : : : : : : : : : : : : : : : : : : : :
313 ATGCCCGAGGAGGGGCGGCAAGTATGTCTCCACGCGCCACAGAGAA 362
||||| : : : : : : : : : : : : : : : : : : : : : : : :
130 ValAlaLysGluValGlyAlaAspAlaValAlaHisGlyCysThrIly 146
||||| : : : : : : : : : : : : : : : : : : : : : : : :
363 GGGGACGATCAGTCAGTCGGTTGAGCTCAGCTCCTACTCAGTGGCCCC 412
||||| : : : : : : : : : : : : : : : : : : : : : : : :
146 sGLYAsnAspGlnValArgPheGluLeuThrPheTYrAlaLeuAsnProG 163
||||| : : : : : : : : : : : : : : : : : : : : : : : :
413 AGATAAGTCATTCCTCCCTGGAGATGCTGAATTCACAAACCGGTTTC 462
||||| : : : : : : : : : : : : : : : : : : : : : : : :
163 IuLeuLysValValAlaProTrpArgGluTrpAsp.....Ile 175
||||| : : : : : : : : : : : : : : : : : : : : : : : :
463 AAGGGCCGCAATGACCTGATGAGTACGCAACAGCACAGGAGTCCCAT 512
||||| : : : : : : : : : : : : : : : : : : : : : : : :
176 ThGlyArgGlnAspAlaIleGluTYrAlaLysLysHisAsnValProVa 192
||||| : : : : : : : : : : : : : : : : : : : : : : : :
513 CCGGTCACATCCCAAGAACCCGTGGAGCATGATGAGAACCTTCATCACA 562
||||| : : : : : : : : : : : : : : : : : : : : : : : :
192 IProValSerLysLysSerIleTYrSerIArgAspArgAsnLeuTrpHis 209
||||| : : : : : : : : : : : : : : : : : : : : : : : :
563 TCAGCTACGAGGCTGGAATCTTGAGAACCCCAAGAACAGGCGCTCCA 612
||||| : : : : : : : : : : : : : : : : : : : : : : : :
209 euseRHisGluGlyAspIleLeuGluAspProAlaAsnGluProLysGlu 225
||||| : : : : : : : : : : : : : : : : : : : : : : : :
613 GGTCCTCAGCAAGAACCCGAGCACCCAGCCCAAGCCCAACCCCTGA 662
||||| : : : : : : : : : : : : : : : : : : : : : : : :
226 AspPheTYrMetSerIleAlaProGluAsnAlaProSerIysProGI 242
||||| : : : : : : : : : : : : : : : : : : : : : : : :
663 CATTCCTGAGATCGATTCACAAAAGGGTCCCTGAGAGTGAACCAAG 712
||||| : : : : : : : : : : : : : : : : : : : : : : : :
242 uTYrLeuGluIleGlyIleIleAlaGlyAlaProValSerIle...Asn 258
||||| : : : : : : : : : : : : : : : : : : : : : : : :
713 TCAAGATGCGACACCACCCACACCTCTGAGACTCTTCATGTCCTG 762
||||| : : : : : : : : : : : : : : : : : : : : : : : :
258 LyrArgAsp.....LeuSerProAlaSerLeuLeuAlaGluLeu 270
||||| : : : : : : : : : : : : : : : : : : : : : : : :
763 AACGAGTCGCGGGAAGCATGGCGTGGGCGCTATTGACATCGTGAGAA 812
||||| : : : : : : : : : : : : : : : : : : : : : : : :
271 AsnGluIleGlyLysHisGlyIleGlyArgIleAspMetValGlnLys 287
||||| : : : : : : : : : : : : : : : : : : : : : : : :
813 CCGCTTCATTGGAATGAGTCCGAGGATATCTACGAGACCCACAGACA 862
||||| : : : : : : : : : : : : : : : : : : : : : : : :
287 nArgLeuVal**MetLysSerArgGlyValTYrGluThrProGIgLYT 304
||||| : : : : : : : : : : : : : : : : : : : : : : : :
863 CCATCTCTTACCATGCTATTAGACATCGAGGCTTCACCATGAGACGG 912
||||| : : : : : : : : : : : : : : : : : : : : : : : :
304 hTrIleMetAlaIleAlaValArgGluLeuGlnSerLeuThrLeuAspArg 320
||||| : : : : : : : : : : : : : : : : : : : : : : : :
913 GAAGTCGCAAAATCAACAAGGCGTGGAGCTGAAATTTGCTGAGCTGTG 962
||||| : : : : : : : : : : : : : : : : : : : : : : : :
321 GluThrMetGlnTrpLysAspIleIleAlaLeuLysTYrAlaGluLeuVa 337
||||| : : : : : : : : : : : : : : : : : : : : : : : :
963 GTATACCGGTTTACGGCTCAGCCCTGAGTGAATTTGCTCCGCACTGCA 1012
||||| : : : : : : : : : : : : : : : : : : : : : : : :
337 ITYrAlaGlyLArgTrpPheAspProLeuArgGlnSerPheAspAlaPhe 354
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1013 TCGCAGAGTCCCAAGGAGCGAGTGAAGGAAATGCGAGTGTCCGTCTC 1062
||||| : : : : : : : : : : : : : : : : : : : : : : : :
354 etGluLysValIThrAlaIThrThrGlySerValIleAlaLysAlaMetVal 370

```

```

1063 AAGGCCAGGTGTCATTCAGTCCGCGGAGTCCCACTGTCTCTACAA 1112
||||| : : : : : : : : : : : : : : : : : : : : : : : :
371 LysGlySerValAsnValAlaSerArgLysSerProTYrSerLeuTYr 387
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1113 TGAGAGCTGTGATGATGAACGTGACGAGTGATATTGAGCCACTGATG 1162
||||| : : : : : : : : : : : : : : : : : : : : : : : :
387 gGluAspLysSerSerPheGluAsnGlyGluIleTYr.AsnGln**Thr 403
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1163 CCACCGGTTTCATCA 1177
||||| : : : : : : : : : : : : : : : : : : : : : : : :
404 ProArgGlySerSer 408

```

```
seq_name: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep:US-09-708-427-55742
```

```
seq_documentation_block:
```

```

; Sequence 55742, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ. ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ. ID NO 55742
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..447
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..447
; OTHER INFORMATION: Ceres Seq. ID 1935961
US-09-708-427-55742

```

```
alignment_scores:
```

```

Quality: 845.50 Length: 389
Ratio: 2.936 Gaps: 6
Percent Similarity: 74.036 Percent Identity: 45.501

```

```
alignment_block:
```

```
US-09-775-693-1 x US-09-708-427-55742 ..
```

```
Align seg 1/1 to: US-09-708-427-55742 from: 1 to: 447
```

```

19 GTGGTTGTGCTACAGTGGCGGCTGGACACCTGTCATCTCGTGTG 68
||| ||||| : : : : : : : : : : : : : : : : : : : : : :
51 Val**LeuAlaTYrSerGIgLYLeuAspHisSerValIleValProTr 67
||||| : : : : : : : : : : : : : : : : : : : : : : : :
69 GCTGAGGAGACAA...GGCTATGACGTCATTCCTATCTGCGCAACAT 115
||||| : : : : : : : : : : : : : : : : : : : : : : : :
67 pLeuArgGluAsnTYrGlyCysGluValValCysPheThrAlaAspAlG 84
||||| : : : : : : : : : : : : : : : : : : : : : : : :
116 GCCAGAGAGAA...GACCTTGAGAGAGCCAGAGAGGACGCTGAGCTT 162
||||| : : : : : : : : : : : : : : : : : : : : : : : :
84 lYgInGlyAlaIleGluLeuGluGlyLeuGluLysLysAlaLysAlaSer 100
||||| : : : : : : : : : : : : : : : : : : : : : : : :
163 GGGGCCAAAAGGTGTCATTCAGATGTCAGGAGGAGTGTGGAGGA 212
||||| : : : : : : : : : : : : : : : : : : : : : : : :
101 GLYAlAcysGlnLeuValValLysAspLeuLysGluGluPheValSerGI 117
||||| : : : : : : : : : : : : : : : : : : : : : : : :
213 GTTCATCTGGCCGGCCATCCAGTCAGCCGACTGTATGAGGACCGCTACC 262
||||| : : : : : : : : : : : : : : : : : : : : : : : :
117 uTYrIleTYrProCysLeuArgIaGlyAlaValTYrGluArgLYSTyTL 134
||||| : : : : : : : : : : : : : : : : : : : : : : : :
263 TCCTGGGACCTCTCTGGCAGCCCTGCATGCGCCGCAACAGAGTGA 312
||||| : : : : : : : : : : : : : : : : : : : : : : : :
134 euleuGlyThrSerMetAlaArgProValIleAlaLysAlaMetValAsp 150

```

```

313 ATGCGCCAGCGGAGGCGGCAAGTATGTCCTCCAGCGGCGGCAAGGAAA 362
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
151 VALALALysGluValGlyAlaAspAlaValAlaHisGlyCysThrGly 167
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
363 GGGGAACGATGAGTCGCGTTTGAGCTGAGCTGCTACTGCGCCGCC 412
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
167 sGlysnAspGlnValAlaPheGluLeuThrPheThrAlaLeuAsnProG 184
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
413 AGATAAGGTCATGCTGCTGCGGAGATGCTGAAATTCACACCGGCTC 462
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
184 luleuysValValAlaProTPrArgGluTrpAsp.....lle 196
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
463 AAGGCGCGCAATGACCTGATGAGTACGCAAGCAACGCGATTCAT 512
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
197 ThrGlyArgGluAspAlaIleGluThrAlaLysHisAsnValProVa 213
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
513 CCCGCTACTCCCAAGAACCCCTGGAGCATGAGTGAAGAACTCATGCA 562
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
213 lProValSerLysLysSerLleTyrSerArgAspArgAsnLeuTrpHisL 230
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
563 TCAGCTACGAGGCTGGAATCCTGGAGAACCCCAAGAACCAACGCGCTCA 612
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
230 euserHisGluGlyAspLleLeuGluAspProAlaAsnGluProLysGlu 246
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
613 GGTCTCTACGCAAGACCCAGACCCCAAGCCCAAGCCCAACCCCTGA 662
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
247 AspMetTyrMetMetSerLleAlaProGluAsnAlaProSerLysProL 263
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
663 CATTCCTGAGATGAGTCAAAAAGGGTCCCTGTAAGGTGACCAAGC 712
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
263 uTyrLeuGluIleGlyLleIleAlaGlyAlaProValSerile...AsnG 279
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
713 TCAGAGATGGCACACCCACCAACACCTCTGGAGCTCTCATGTCAC 762
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
279 LyrArgSp.....LeuSerProAlaSerLeuLeuAlaGluLeu 291
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
763 AAGCAAGTCGCGGCAAGATGCGCTGGCGCTATTGACATCGTGGAGA 812
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
292 AsnGluIleGlyLysHisGlyIleGlyArgIleAspMetValGluAs 308
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
813 CCGCTTCATTGGAATGATCCGAGGATATCTCGAGACCCCAAGCGCA 862
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
308 nArgLeuVal**MetLysSerArgGlyValTyrGluThrProGlyLys 325
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
863 CCAATCCCTTACCATGCTATTGACATCGAGGCTTCACCATGACCGG 912
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
325 hTrIleMetAlaAlaAlaValAlaArgIleuGluSerLeuThrLeuAspArg 341
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
913 GAAGTGGCAAAATCAACAAGCGCTGGCTGAATTTGCTGAGCTGT 962
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
342 GluThrMetGlnTrpLysAspLleIleAlaLeuLysTyrAlaGluLeuVa 358
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
963 GTATACCGGTTTACGCGCTGAGCTGATGATTTGTCCGCCACATGCA 1012
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
358 lTyrAlaGlyArgTrpPheAspProLeuArgGlnSerPheAspAlaPheM 375
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1013 TCGCCAAGTCCACAGAGCGAGTGAAGCAAGTGCAGTGTCCGCTCT 1062
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
375 etGluLysValThrAlaThrThrThrLysSerValThrLeuLysLeuTyr 391
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1063 AAGGCGCAGGTGATCATCTCGCGCGGAGAGTCCCACTGCTCTCAAA 1112
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
392 LysGlySerValAsnValAlaSerArgLysSerProTyrSerLeuTyrAr 408
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1113 TGAGAGCTGCTGAGCATGAACGTGCAAGGCGATTAAGCCACATCATG 1162
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
408 gGluAspLleSerSerPheGluAsnGlyGluLleTyr.AsnGln**Thr 424
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1163 CCACCGGGTTCATCA 1177
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
425 ProArgGlySerSer 429
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::

```

seq_name: /cgn2_6/plotdata/2/paa/US09_NEW_COMB pep: US-09-897-516-5764

```

seq documentation block:
; Sequence 5764, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Cordin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spilidonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 5764
; LENGTH: 410
; TYPE: PR
; ORGANISM: Xenorhabdus sp.
US-09-897-516-5764

```

```

alignment_scores:
  Quality: 819.50      Length: 405
  Ratio: 2.836         Gaps: 6
  Percent Similarity: 71.358      Percent Identity: 45.185

```

alignment_block:

US-09-775-693-1 x US-09-897-516-5764 ..

Align seg 1/1 to: US-09-897-516-5764 from: 1 to: 410

```

19 GTGGTTTCGCTTACAGTGGCGGCTGAGACACCTGTCATCTCGTGTG 68
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
15 lIleValleuAlaTyrSerLcylGlyLeuAspTrpSerAlaIleIleProTr 31
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
69 GCTGAAGGAACA.....GGCTATGACGTCATTCCTATCTGGCCANCA 112
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
31 pLeuysGluHisTyrGlyAsnGlyAspValIleAlaPheValAlaAspV 48
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
113 TTGGCCAG...AAGAAAGCTTCGAGAGAACCCAGAAAGAGCACTGAG 159
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
48 aGlyGlnSerArgGlnAspLeuGluGlyIleGluGlnLysAlaLeuArg 64
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
160 CTTGGGCGCCAAAAGGTGTCATTGAGGATGTCACGAGGAGATTGTGGA 209
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
65 SerGlyAlaSerLcylGlyHisIleAlaAspLeuArgGluGluPheIle 81
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
210 GGAGTCATCTGCGCGCCATCCAGTCCAGCGCACTGATGAGAGCGCT 259
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
81 sGluTyrValTyrProValLeuLysThrGlyAlaLeuTyrGluGlySerT 98
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
260 ACCTCTGGCACTCTCTTGGCAGGCGCTGCATGCGCCGCAAAAGTGT 309
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
98 yTrLeuLeuGlyThrSerMetAlaArgProIleIleAlaLysAlaGlnVal 114
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
310 GAATGCGCCAGCGGAGGCGGCAAGTATGTCCTCCAGCGGCGCAAGG 359
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
115 GluLeuAlaLeuLysValAlaLysAlaAspAlaLeuAlaHisGlyAlaThrCl 131
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
360 AAGGGAACGATCAGTCCGCTTGGAGTCAAGCTGCTACTGAGCGCC 409
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
131 yLysGlyAsnAspGlnValAlaPheGlnSerThrTyrThrAlaLeuAlaP 148
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
410 CCCAGATAAAGTTCATCTCTGCTGCGAGATGCTGCTCAATTATCAACCGG 459
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
148 roHisLeuLysValAlaAlaProTPrArgGluTrpAsp..... 160
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
460 TTCAGGCGCGCAATGACCTGATGAGTACGCAAAACACACGCGGATTC 509
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

161 LeuArgSerArgLalLeuLeuAspTyrLeuLysValAlaArgSpLier 177
510 CATCCCGGTCACCTCCAGAGACCCGTGGAGCATGATGAGAACCTTCATGC 559
177 othrThrAlaThrLeuGluLysLleTyrSerArgAspGluAsnAlaIrrh 194
560 ACATGACGACGAGGCTGGAATCCTGAGACACCCCAAGAACCAAGGCGCT 609
194 lslSerThrGluGlyGlyValLeuGluAsnThrTrpAsnAlaAlaAsn 210
610 CCAGGCTCTACACGAGACCCAGACCCCAAGCCCAAGACCCCAAGACCC 659
211 LysAspCysTrpGluTrpThrValGluProGluAspAlaProAspGluPr 227
660 TGCAATTCCTGAGATCGAATCAAAAAGGGTCCCTGTGAAGGTGACCA 709
227 oGluTyrLeuSerValThrValGluLysGlyValValGlyVal...A 243
710 AGCTGAGAGATGACACACCCACACACCTCCTGGAGCTTCATGTAC 759
243 snGlyLysIleLeuSerProTyrGlnCysLeuAsnGlu..... 255
760 CTGAAGAGAGTCCGCGGACAGCATGGCGGCGCTATTCATGATGATGCA 809
256 LeuAsnGluLeuGlyAlaLysHisGlyIleGlyArgIleAspIleValGI 272
810 GAACGCGCTCATTTGAATGAATGCCGAGATCTACGACGACCCACAG 859
272 uAsnArgLeuValGlyMetLysSerArgLysCysTyrSerIuThrProGlyG 289
860 GCACCATCCTTACCATCTCATTTAGACATGAGAGCCCTTCACCATGAGC 909
289 lYThrIleMetMetAlaLeuArgGlyIleGluGlnLeuValLeuAsp 305
910 GCGGAGAGGCGCAAAATCAACAGAGCCTGGGCTTGAATTTCTGAGACT 959
306 ArgAspSerPheLysTrpArgGlnIleLeuGlyLeuGluMetSerTyrVa 322
960 GGTGTATACCGGTTTACGCGCTAGCCCTGAGATGTAATTTGTCCGCCACT 1009
322 lValTyrAspGlyArgTrpPheValProLeuArgGlnSerIleGlnAla 339
1010 GCATGCGCAAGTCCAGAGAGGAGTGAAGGAAAGTGCAGGTCCGTC 1059
339 lAlaGluMetLeuAlaGluAsnValSerGlyGluValIleLeuLysLeu 355
1060 CTCAAGGCGAGGTGTACATCTCGGCGGAGTCCACACTGCTCTCTA 1109
356 TyrLysGlyGluValThrAlaThrGlnLysThrSerProHisSerLeuTy 372
1110 CAATGAGAGAGCTGGTGAAGATGAAGCTGAGATGATGAGACCACTG 1159
372 rSerGluGluPheAlaThrPheGlyGluAspGluValTyrAspHisSerh 389
1160 ATGCCACCGGTTTCATCACTCAATTCCTGAGGCTGAAGAAATATCAT 1209
389 lAlaGluGlyGluPheIleArgLeuHisSerLeuSer...SerArgIleArg 404
1210 CGTCTCCAGAGCAAG 1224
405 AlaLeuLysSerLys 409

seq_name: /cgn2_6/ptodata/2/paa/US09_NFM_COMB pep: US-09-708-427-11258

seq_documentation_block:
; Sequence 11258, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427

```

```

; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11258
; LENGTH: 498
; TYPE: PRF
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..498
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..498
; OTHER INFORMATION: Ceres Seq. ID 1822815
; US-09-708-427-11258

alignment_scores:
Quality: 815.00 Length: 396
Ratio: 2.890 Gaps: 5
Percent Similarity: 71.212 Percent Identity: 42.172

alignment_block:
US-09-775-693-1 x US-09-708-427-11258 ..
Align seq 1/1 to: US-09-708-427-11258 from: 1 to: 498

19 GTGGTCTGTGCTACAGTGGCGGCTGACACCTGTCATCCGTGTCG 68
128 ValValLeuAlaTyrSerGlyGlyLeuAspThrSerValIleValProTr 144
69 GGTGAAGAGCAAA...GGCTATACGTCATTCGCTATTCGCCACATTG 115
144 PLeuLysGluAsnTyrGlyCysGluValValCysPheThrAlaAspValG 161
116 GCCAGAG...GAAGACTTGAGAGAGCCAGAGCAAGAGCAGTGAAGCTT 162
161 LysGlnGlyIleLysGluLeuGluGlyLeuGluGlnLysAlaLysAlaSer 177
163 GGGGCAAAAGAGTTCATTCATGAGATGTCAGCAGGAGTGTGTGAGGA 212
178 GlyAlaSerGlnLeuValAlaLysAspLeuThrGluGluPheValLysAs 194
213 GTTCATCTGGCGCGCATCCAGTCAGCGGCACTGTATGTAGACCCGTCAC 262
194 PheIlePheProCysLeuArgAlaGlyAlaIleTyrGluArgLysTyrL 211
263 TCCTGGCACTCTCTTGCAGAGCCCTGCATCGCCGCAAAACAGTGA 312
211 euLeuGlyThrSerMetAlaArgProValIleAlaLys... 223
313 ATGCCACCGAGGAGGCGGCAAGTATGTGTCCACGCGGCCACAGAA 362
223 ..... 223
363 GGGGAGCATGACGTCGGTTGAGTCAGTGTACTACATGCGCCGCC 412
224 .....ValArgPheGluLeuThrPhePheSerLeuAsnProG 236
413 AGATAAAGGTCAATTCCTCCCTGAGAGATGCCCTGATTCACAAACGGTTC 462
236 lueLysValValAlaProTrpArgGluTrpGlu.....Ile 248
463 AAGGCGCGCAATGACCTGATGAGTACGCAAGCAACACGAGATCCCAT 512
249 GlnGlyArgGluAspAlaIleGluTyrAlaLysLysHisAsnValProVa 265
513 CCCGCTCACTCCCAAGAACCCGTGAGCATGATGAGAACCTGACCA 562
265 lProValThrLysLysSerIleTyrSerArgAspArgAsnLeuTrpHisL 282
563 TCAGCTAGAGGCTGGAATCTCGAGAACCCCAAGCAACCAAGCCCTTCA 612
.....

```



```

1 ORGANISM: Aspergillus fumigatus
US-09-675-784A-12974

Alignment_scores:
    Quality: 759.00      Length: 233
    Ratio: 4.059         Gaps: 0
    Percent Similarity: 80.258      Percent Identity: 60.944

Alignment_block:
US-09-775-693-1 x US-09-675-784A-12974  ..

Align seg 1/1 to: US-09-675-784A-12974 from: 1 to: 257

7 ACCAAGGCTCCGTGTTCTGACCTACGTGCGGCTGACACCTGCTG 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 SerLysgLySvalcYsLeuAlaPheSerlgYlLeuAsPThrSera 38
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 CATCTGCTGTGCTGAAGACAGGCTATGACGTCACTGATCTATCG 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 lLleuLysTrPleuLleAspgLugLyTrglValAlaPhePmra 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 CCACACTGGCCAGAGAGAACTCTGAGAAACCAGAAAGGACTGTG 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 lAsnValGlYglInclunLAsPheAsPAlleargLgluLysAlleu 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 AAGCTTGGGGCCAAAAGTTCATTGAGAGTACGAGGAGACTTGT 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 LysLeuGlYAlaValLyscYsGluLleAlaAsPLeuArgYgluPhe 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 GGAGGAGTCATCTGGCCGCCCATCTCCAGTCCAGCACTGTATGAGACC 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 eglInLleuCySphProAlaIleAlaCySAsnAlaIleTyrgLuAsnV 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
257 GCTACCTCTCTGGGACACTCTCTGCTCCAGGCGCTGCATCGCCGCAACAA 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 alTyrgLeuLengLyThrSerLeuAlaArgProValIleAlaArgAlaGln 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 GTGGAATCGCCCGACGGAGGGGGCCAAATGATGTGCCACGGCGCAC 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 lLAspValAlaLysTrgluLelcYsPheAlaValSerIsgLySph 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 AGGAAAGGGGAAAGATAGAGTCCGTTGAGCTCAGCTGCTACTG 406
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 rgLySglYAsnAspGlnValArgPheGluLeuAlaPheTyrgAlaLeuG 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
407 CCCCCAGATAAGGTCATGCTGCTCCCTGGAGAGTCCCTGAATTTCAAC 456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 lnpProAsPLeuLysValIleAlaIProTrpArgAsPProAlaPheTyrgAsn 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
457 CGGTCAAGGGCCGCAATGACCTGATGGAGTACGCCAAAGACAGACGGGAT 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 ArgPheAlaGlYargAsnAsPLeuSerTyrgAlaAlaGluLySgYl 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
507 TCCCATCCCGGTCACTCCCAAGACCCGTGGAGCATGGATGAGAGACTCA 556
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 eProValThrSerThrLysAlaLysTrPTrSerMetCAsPgluAsnLeuA 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 TGCACTACGCTACGAGGCTGGAATCTGGAGAAGCCCAAGACCAAGCG 606
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 lAhIscYSerTyrgluAlaGlylLeuLengLuAsPProAsnValThrPro 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
607 CCTCCAGGTCTTACACGAAGACCCAGGACCCAGCCAAAGCCCCACAC 656
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 ProAlaAspMetTyrgLysLeuThrGlnAsPProLeuThrAlaIProAsPgl 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
657 CCTGACATCTCTCAGATCGAGTTCCAAAAAGGGGTCCCTGTGAAGGTG 705
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 nProclunAsPheThrValIshPheGluThrGlylLeProValLysLeu 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: /cgn2_6/ptodata/2/paa/US09_NMW_COMB.pep:US-09-708-427-79359
seq_documentation_block:

```

```

; Sequence 79359, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708, 427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79359
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..308
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..308
; OTHER INFORMATION: Ceres Seq. ID 1965190
US-09-708-427-79359

```

```

alignment_scores:
  Quality: 655.50      Length: 305
  Ratio: 2.953         Gaps: 3
  Percent Similarity: 72.787   Percent Identity: 44.590

```

alignment_block:

```
US-09-775-693-1 x US-09-708-427-79359 ..
```

```
Align seg 1/1 to: US-09-708-427-79359 from: 1 to: 308
```

```

277 CTTGCCAGGCCCTGATCGCCGCAACAAAGTGGAAATCGCCGAGGGGGA 326
    :::::::::::::::::::::
  1 MetAlaArgProValIleAlaIysAlaMetValAspValAlaIysGluVal 17
327 GGGGGCCCAAGTATGTGTCGCCAGCGCCGACAGAGAAAGGGAGCAATGAG 376
    |||||
  17 IGIAlaAspAlaValAlaIleHisGlyCysThrGlyIysGlyAlaMetGluVal 34
377 TCCGGTTGAGCTGAGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 426
    |||||
  34 AlaArgPheGluLeuThrPheThrAlaLeuAsnProGluLeuIysValVal 50
427 GCTCCCTGGAGAGATGCTCTCAATCTTACAACCGGTTCAAGGCGCCGATGA 476
    |||||
  51 AlaProThrPargGluTrpAsp.....IleThrGlyArgGluAsn 63
477 CCTGATGAGTACGCAAGCAACGAGGATCCCATCCCGGTCACTCCCA 526
    |
  63 PAlaIleGluTrAlaIysIshisAsnValProValProValSerLysL 80
527 AGAACCCGTGAGCATGATGAGAACCTCATCATCATCATCATCATCATCATG 576
    ||:::
  80 YSerIleIerYserArgAspArgAsnLeuTrpHisLeuSerHisGluGly 96
577 GGAATCTTGAGAACCCCAAGAACGAGCGCTCCAGGTCTCTACACGAA 626
    |||||
  97 AspIleLeuGluAspProAlaAsnGluProLysGluAspMetLysMele 113
627 GACCCGAGACCCGAGCAAGCCGCAACACCCGTCGATCTTCGAGATCG 676
    ::::
  113 tSerIleAlaProGluAsnAlaProSerLysProGluIleuGluIleG 130
677 AGTTCAAAAAAGGGTCCCTGAGAGGTGACCAAGTCAGATGAGGACCC 726
    ::::
  130 IyIleIleAlaGlyAlProValSerIle...AsnGlyArgAsp..... 143
727 ACCCAACAGACCTCTCTGAGCTCTTCATGTACTGAAAGAGTGGCGGG 776
    ::::
  144 .....LeuSerProAlaSerLeuLeuAlaGluLeuAsnGluIleGlyG 158

```

```

777 CAAGCATGGGCGGGCGGATTTGACATGTCGAGAACCCGCTTCATTTGAA 826
    |||||
  158 YLrshHisGlyIleGlyArgIleAspMetValGluAsnArgLeuValCysM 175
827 TGAAGTCCGAGGATTCACAGAGACCCGAGGAGGACCATCTTACCAT 876
    |||||
  175 eLysSerArgGlyAlaTyrgluThrProGlyIleThrIleMetAlaAla 191
877 GCTCATTTGACATGAGCGCTTCACATGAGACCGGAGAGTGGCAAAAT 926
    |||
  192 AlaValArgIleuLeuSerLeuThrLeuAspArgGluTrpMetGlnTr 208
927 CAACAGAGCCGCGGCTTGAATTTGCTGAGCTGCTGATATACGGTTTAC 976
    ||:::
  208 pLysAspIleIleAlaLeuLysTyrgluGluLeuValTyrgluArgT 225
977 GGCCTAGCCCTGAGTGTGAATTTGTCGCCACATCGCCCAAGTCCAG 1026
    ::::
  225 rPheAspProLeuArgGlnSerPheAspAlaPheMetGluLysValThr 241
1027 GAGCGATGAGAGGCAAGTGCAGGTGCGGCTCTCAAGGCGCAGGTGA 1076
    ::::
  242 AlaThrThrThrGlySerValThrLeuLysLeuIyrgLysIserValAs 258
1077 CATCTGCGCGGAGAGTCCCATGCTCTCTCAATGAGAGACTGTGTA 1126
    ::::
  258 nValAlaSerArgLysSerProIySerLeuTyrgluAspIleSerS 275
1127 GCATGACGTGCAGGTGATATGAGCCATGATGACACCGGCTTCATC 1176
    ||:::
  275 erPheGluAsnGlyGluIleTyrgAsnGlnAlaAspAlaGlyPheIle 291
1177 AACATCAATTCCTTC 1191
    ::::
  292 ArgLeuTyrgLysLeu 296

```

```
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW.COMB.pep:US-09-708-427-55744
```

```

seq_documentation_block:
; Sequence 55744, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708, 427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55744
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..309
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..309
; OTHER INFORMATION: Ceres Seq. ID 1935963
US-09-708-427-55744

```

```

alignment_scores:
  Quality: 639.50      Length: 301
  Ratio: 2.947         Gaps: 4
  Percent Similarity: 72.093   Percent Identity: 44.850

```

alignment_block:

```
US-09-775-693-1 x US-09-708-427-55744 ..
```

```
Align seg 1/1 to: US-09-708-427-55744 from: 1 to: 309
```

```

277 CTTGCCAGGCGCCCTGCATCGCCCGCAACAGTGAATCGCCAGCGGGA 326
    ::::::::::::::::::::
1 MetalAaArgProValIleAlaLysAlaMetValAspValAlaLysGluVal 17
327 GGGGGCCAAAGTATGTGTCCACGGCGCCACAGAAAGGGAACATCGAG 376
    ::::::::::::::::::::
17 LcylAlaSprAlaValAlaIleHisGlyCysThrGlyLysGlyAsnAspGln 34
377 TCCGGTTTGAGCTCAGCTGCTACGTCAGCGCCCGCCAGATTAAGTCAAT 426
    ::::::::::::::::::::
34 AlaArgPheGluLeuThrPheTyrAlaLeuAsnProGluLeuLysValVal 50
427 GCTCCCTGGAGAGTGCCTGAATTCTACAAACCGGTTCAAGGGCGCCAGTGA 476
    ::::::::::::::::::::
51 AlaProTTPArgGluTTPasp.....IleThrGlyArgGluAs 63
477 CCTGATGAGTACGCAACAGCAACAGCGGATTCCTCCCGGTCACTCCCA 526
    ::::::::::::::::::::
63 PalAlaIleGluTyrAlaLysLysHisAsnValProValProValSerLysL 80
527 AGAACCCCGGAGCATGATGAGAACCTTCATGCACATGACGATGACGAGGCT 576
    ::::::::::::::::::::
80 ySerLleTyrSerArgAspArgAsnLeuTTPHisLeuSerHisGluGly 96
577 GGAATCCTGAGAGACCCCAAGCAACAGCGCCCTCCAGGCTCTACACGAA 626
    ::::::::::::::::::::
97 AspIleLeuGluAspProAlaAsnGluProLysGluAspMetLysMetMe 113
627 GACCCAGGACCCCAAGACCCCGCCACACCGCTGACATTCCTCGAGATCG 676
    ::::::::::::::::::::
113 tSerIleAlaProGluAsnAlaProSerLysProGluTyrLeuGluIleG 130
677 AGTTCAAAAAGGGGTCCTGTAAGGTGACCAACGTCAGTGAAGATGGACAC 726
    ::::::::::::::::::::
130 LylleIleAlaGlyValProValSerLys...AsnGlyArgAsp..... 143
727 ACCCAGCAGACCTCTTGAGGCTTCATGCTGACGTAACGAAAGTCGCGGG 776
    ::::::::::::::::::::
144 .....LeuSerProAlaSerLeuLeuAlaGluLeuAsnGluIleGlyL 158
777 CAAGCATGGCGTGCGCGCTATTGACATCGTGGAGAACCGCTTCATTGAA 826
    ::::::::::::::::::::
158 LysHisGlyIleGlyArgIleAspMetValGluAsnArgLeuVal***M 175
827 TGAAGTCGGAGGTATCTACGAGACCCAGCGGCGCATCCCTTTCAT 876
    ::::::::::::::::::::
175 eLysSerArgIlyValTyrGluThrProGlyGlyThrIleMetAlaIa 191
877 GCTCATTTAGACATCGAGGCTTCACCATGACCGGGAAGTGGCAAAAT 926
    ::::::::::::::::::::
192 AlaValArgGluLeuGluSerLeuThrLeuAspArgGluThrMetGlnTr 208
927 CAACAACAGCGCTGGCTGAATTTGCTGAGCGTGGTGTATACCGGTTTAC 976
    ::::::::::::::::::::
208 pLysAspIleIleAlaLeuLysTyrAlaGluLeuValTyrAlaGlyArgT 225
977 GGCCTAGCCCTGAGTGAATTTGTCGCGCACATCCGCAAGTCCCGAC 1026
    ::::::::::::::::::::
225 rPheAspProLeuArgGlnSerPheAspAlaPheMetGluLysValThr 241
1027 GACCGAGTGAAGGAAAGTGAAGTGTCCGTCCTCAAGGGCGGAGGTGTA 1076
    ::::::::::::::::::::
242 AlaThrThrThrGlySerValThrLeuLysLeuTyrLysGlySerValAs 258
1077 CATTCCTGGCGGGAGTCCCGACTGCTCTCATCAATGAGGAGGATGCTGA 1126
    ::::::::::::::::::::
258 nValAlaSerArgLysSerProTyrSerLeuTyrArgGluAspIleSerS 275
1127 GCATGACAGCTGCAGGGTATATGAGCAACTGATGCACACGGGTTTCAT 1176
    ::::::::::::::::::::
275 eRheGluAsnGlyGluIleTyr.AsnGln***ThrProArgGlyLysSer 291

```

```

1177 A 1177
291 r 291
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-708-427-11259

```

```

seq_documentation_block:
; Sequence 11259, Application US/09708427

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: N. ALEXANDROV et al.

```

```

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID

```

```

; TITLE OF INVENTION: THEREBY

```

```

; FILE REFERENCE: 2750-1243P

```

```

; CURRENT APPLICATION NUMBER: US/09/708,427

```

```

; NUMBER OF SEQ ID NOS: 85364

```

```

; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 11259

```

```

; LENGTH: 283

```

```

; TYPE: PRT

```

```

; ORGANISM: Arabidopsis thaliana

```

```

; FEATURE:

```

```

; NAME/KEY: misc.feature

```

```

; LOCATION: 1..283

```

```

; NAME/KEY: misc.feature

```

```

; LOCATION: 1..283

```

```

; OTHER INFORMATION: Ceres Seq. ID 1822816

```

```

US-09-708-427-11259

```

```

align_scores:
Quality: 598.00 Length: 308
Ratio: 2.875 Gaps: 3
Percent Similarity: 67.532 Percent Identity: 40.260

```

```

alignment_block:
US-09-775-693-1 x US-09-708-427-11259 ..

```

```

Align seg 1/1 to: US-09-708-427-11259 from: 1 to: 283

```

```

277 CTTGCCAGGCGCCCTGCATCGCCCGCAACAGTGAATCGCCAGCGGGA 326
    ::::::::::::::::::::
1 MetalAaArgProValIleAlaLysAlaMetValAspValAlaLysGluVal 8
327 GGGGGCCAAAGTATGTGTCCACGGCGCCACAGAAAGGGAACATCGAG 376
    ::::::::::::::::::::
9 .....V 9
377 TCCGGTTTGAGCTCAGCTGCTACGTCAGCGCCCGCCAGATTAAGTCAAT 426
    ::::::::::::::::::::
9 AlaArgPheGluLeuThrPhePheSerLeuAsnProGluLeuLysValVal 25
427 GCTCCCTGGAGAGTGCCTGAATTCTACAAACCGGTTCAAGGGCGCCAGTGA 476
    ::::::::::::::::::::
26 AlaProTTPArgGluTTPGln.....IleGlnGlyArgGluAs 38
477 CCTGATGAGTACGCAACAGCAACAGCGGATTCCTCCCGGTCACTCCCA 526
    ::::::::::::::::::::
38 PalAlaIleGluTyrAlaLysLysHisAsnValProValProValThrLysL 55
527 AGAACCCCGGAGCATGATGAGAACCTTCATGCAATGAGGAGGATGCTGA 576
    ::::::::::::::::::::
55 ySerLleTyrSerArgAspArgAsnLeuTTPHisLeuSerHisGluGly 71
577 GGAATCCTGAGAGACCCCAAGCAACAGCGCCCTCCAGGCTCTACACGAA 626
    ::::::::::::::::::::
72 AspIleLeuGluAspProAlaAsnGluProLysGluAspMetLysMetMe 88
627 GACCCAGGACCCCAAGACCCCGCCACACCGCTGACATTCCTCGAGATCG 676
    ::::::::::::::::::::
88 tSerValAspProGluAspProAlaProAspGlnProGluTyrIleGluIleG 105

```

```

677 AGTCAAAAAGGGCTCCCTGTGAGGTGACCAAGCTCAAGATGCGACC 726
      ::::::::::::::::::::
105 LylleGIuserGIyleuProValAlaleuasnGly..... 116
      ::::::::::::::::::::
727 ACCCAGACGCTCCTTGAGCTCTTCATGTACCTAGACGAGTCCGGG 776
      ::::::::::::::::::::
117 LysAlaleuSerProAlathrLeuLeuAlaGluLeuAsnThrLleGI 133
      ::::::::::::::::::::
777 CAAGCATGGCGCTGGCCGTATTGACATCGTGAGAACCGCTTCATTTG 826
      ::::::::::::::::::::
133 YLysHISGlyIleGlyArgIleAspMetValGluAsnArgyleValGly 150
      ::::::::::::::::::::
827 TGAAGTCCGAGGTATCTACGAGACCCAGCAGGACCATCTTACAT 876
      ::::::::::::::::::::
150 etlYSerIArgGlyValTYrGluThrProGIyGIYThrIleuPheAla 166
      ::::::::::::::::::::
877 GCTCATTTAGACATCGAGGCGCTTACCACCGGAGGAGTGGCCAAAT 926
      ::::::::::::::::::::
167 AlValAlGInGluLeuGIuserLeuThrLeuAspArgGluSerIleG 183
      ::::::::::::::::::::
927 CAACAAGGCGCTGGCCGTGAAATTTGCTGAGCTGTGTATACCGGTT 976
      ::::::::::::::::::::
183 llysAspThrLeuAlaleuLysTYrAlaGluMetValTYrAlaGlyArg 200
      ::::::::::::::::::::
977 GGGCTAGCCCTGAGTGTGAATTTGTCCGCCACTGCATGCCCAAGTCC 1026
      ::::::::::::::::::::
200 rPheAspProLeuArgGluSerMetAlaPheMetGluLysIleThr 216
      ::::::::::::::::::::
1027 GAGCGAGTGAAGGCAAGTGCAGGCTGCCGCTCAAGGCGCAGGTGA 1076
      ::::::::::::::::::::
217 GluThrThrThrGlySerValThrLeuLysLeuTYrLysGlySerVal 233
      ::::::::::::::::::::
1077 CATCTCGGCGCGGAGTCCCACTGCTCTTACATGAGAGAGAGTGTGA 1126
      ::::::::::::::::::::
233 rValThrGlyArgGluSerProAsnSerLeuTYrArgGluAspIleSer 250
      ::::::::::::::::::::
1127 GATGAGCTGCAAGGTGATTTAGACCCACTGATGCCACCGGCTTATC 1176
      ::::::::::::::::::::
250 erPheGluGIYserGIuIleTYrAsnGluAlaAspAlaIleGlyPhe 266
      ::::::::::::::::::::
1177 AACATCAATTCCCTCAGCTGAAG 1200
      ::::::::::::::::::::
267 ArgLeuTYrGIYleuProMetLys 274
      ::::::::::::::::::::

seq_name: /cgn2_6/plodata/2/paa/us10_NEW_COMB.pep:us-10-029-386-32991

seq_documentation_block:
: Sequence 32991, Application US/10029386
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
: FILE REFERENCE: AROMICA-X-2
: CURRENT APPLICATION NUMBER: US/10/029,386
: NUMBER OF SEQ ID NOS: 34288
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 32991
: LENGTH: 138
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC004616.1
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.83
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
: OTHER INFORMATION: SWISSPROT HIT: P00966, EVALU6 6.00e-48
US-10-029-386-32991

```

```

alignment_scores:
  Quality: 477.00      Length: 110
  Ratio: 4.500        Gaps: 0
  Percent Similarity: 96.364      Percent Identity: 84.545

```

```

alignment_block:
US-09-775-693-1 x US-10-029-386-32991 ..

```

```

Align seg 1/1 to: US-10-029-386-32991 from: 1 to: 138

```

```

1 ATGTCACCAAAAGGCTCCGTTGCTTGGCTTACAGTGGCGGCTGACAC 50
  ::::::::::::::::::::
28 ValSerSerLysGlySerMetValleuAlaHisSerGlyLeuAspThr 44
  ::::::::::::::::::::
51 CTGCGCATCCCTCGTGGCGTGAAGCAAGAGCTATGAGCTCATTTGCT 100
  ::::::::::::::::::::
44 rSerCysIleLeuMetCitrLeuLysGluGlnGlyHisAspValIleAla 61
  ::::::::::::::::::::
101 ATCTGGCCCAACATTGGCCAGAGAGACCTTCAGAGAACCCAGAGAGAG 150
  ::::::::::::::::::::
61 YrleuAlaLysIleSerGlnLysGluAspPheGluAspArgLysLys 77
  ::::::::::::::::::::
151 GCACCTGAAGCTTGGGCGCAAAAGGTGTTCATTGAGATGTACAGCAGGA 200
  ::::::::::::::::::::
78 AlatrPheLysGluGlyAlaLysLysValPheIleAspAspValSerArg 94
  ::::::::::::::::::::
201 GTTTGTGAGAGAGTTCATCTGCGCGGCGCATCCAGTCCAGGCACTGATG 250
  ::::::::::::::::::::
94 PheValAlaGluIlePheIleThrProAlaIleGlnSerSerIleLeuTYr 111
  ::::::::::::::::::::
251 AGGACCGCTACCTCTGAGGACCTCTCTTGCAGAGGCGCTCATGCGCCG 300
  ::::::::::::::::::::
111 lAspTYrTYrLeuLeuGlyTYrSerLeuAlaArgProCysIleAlaHis 127
  ::::::::::::::::::::
301 AAACAAGTGAATGCCGCCAGCGGAGGG 330
  ::::::::::::::::::::
128 LysGlnValGluIleThrGlnGlnGluGly 137
  ::::::::::::::::::::

```

Wed Feb 13 07:36:01 2002

us-09-775-693-1.rapn

Page 15

